FOLLOWING SPOKEN INSTRUCTIONS IN L1 AND L2

FOLLOWING SPOKEN INSTRUCTIONS IN L1 AND L2: THE EFFECTS OF LANGUAGE DOMINANCE, WORKING MEMORY, COGNITIVE COMPLEXITY, AND LANGUAGE ACQUISITION BACKGROUND

BY

EDALAT SHEKARI, M.A., B.A.

A THESIS SUBMITED TO
THE DEPARTMENT OF LINGUISTICS & LANGUAGES
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENT
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

McMaster University

Doctor of Philosophy (2020) McMaster University
Linguistics & Languages Hamilton, Ontario, Canada

TITLE: Following Spoken Instructions in L1 and L2: The Effects of Language Dominance, Working Memory, Cognitive Complexity, and Language Acquisition Background

AUTHOR: Edalat Shekari
M.A., B.A. Linguistics & Applied Linguistics

SUPERVISOR: Dr. Elisabet Service

NUMBER OF PAGES: xxviii, 243

To my dear wife, Afsaneh,

and my beloved children, Saeid & Melika

Abstract

This thesis investigates cognitive aspects of following spoken instructions in bilinguals' dominant and non-dominant languages. The aim was to increase our understanding of the cognitive benefits and costs of information processing and performance in a second language (L2) compared with a first language (L1). I hypothesized that L2 tasks are associated with more computational and cognitive costs, even in highly proficient bilinguals. In three empirical studies, I examined to what extent factors such as the language of task, working memory capacity, phonological memory, and variables in bilinguals' language history affect processing and performance of sequences of oral instructions in L1 and L2. Furthermore, I manipulated the psycholinguistic complexity of sequential temporal order to see how it influences bilinguals' processing and performance, and how it interacts with other variables. Contrary to similar monolingual studies, I tested bilingual participants in two separate sessions, and applied mixed effects logistic regression models to analyze the data. The results suggest that language dominance significantly affects bilinguals' processing and performance of sequential verbal instructions, with a disadvantage for tasks presented in the non-dominant language. Bilinguals' recall accuracy was consistently superior when the target sentences were presented in L1. Working memory capacity and phonological memory correlated with instruction-following abilities, especially, in less-proficient bilinguals. Individuals with more available working memory resources were more likely to have better processing and performance in following sequences of spoken instructions. The psycholinguistic complexity of temporal order expressions affected the ability to follow sequential oral instructions, causing lower recall accuracy when the surface order of events was incongruent with the factual order of events in complex instructions. A number of variables in bilinguals' language background, specifically the age of initial L2 acquisition, level of L2 proficiency, L2 use, and L2 exposure also predicted performance, especially, in less-skilled bilinguals. To the best of my knowledge, this is the first empirical study that has investigated the effects of using a non-dominant language on following instructions modelled on real-world tasks in an office workplace.

Acknowledgements

This thesis would not have been possible without the help of some faculty members and my fellow students in the Department of Linguistics and Languages, the Language, Memory, and Brain Lab (LMBLab), and the centre for Advanced Research in Experimental and Applied Linguistics (ARiEAL) at McMaster University.

First and foremost, I would like to thank my supervisor, Elisabet (Lili) Service, who never hesitated to support me during my studies. I was privileged to be one of Lili's Ph.D. students. She agreed to take my supervisory responsibilities when I applied to McMaster University although we never met each other in person. She patiently and professionally guided me throughout my graduate studies. We spent hours and hours on designing my experiments or modifying them once a new idea emerged. Her superior empirical research skills, patience, and unique personal manners were all incentives for a student who was new and naïve to the empirical research and Neurolinguistics. I will always be grateful for Lili's academic and non-academic supports and her kind-heartedness. She was well-aware of the needs of a mature student with dependents and were mindful so that my studies would not affect my family. Also, I appreciate her funding, some of it enabled us to pay bilingual participants for their participation in bilingual experiments. Here, I would like to thank her for giving so much of her time and energy. All her efforts yielded fruitful results and I have become a better linguist and a researcher.

Secondly, I would like to thank the members of my (internal) dissertation supervisory committee, Karin Humphreys and Scott Watter. Karin and Scott joined my dissertation committee in year three of my studies. Both are inspiring professors and mentors, and I am grateful for their helpful feedback during running the experiments, analyzing the data and compiling my thesis. They put a lot of efforts to read the chapters of my thesis and provided me with constructive feedback. I would like to thank Anna Moro for being my co-supervisor in the first year of my studies. Unfortunately, we missed Anna in the supervisory committee because of her duties and assignments as the associate dean of humanities. Also, I would like to thank Ivona Kucerova and Arsalan Kahnemuyipour (University of Toronto) for their efforts as committee members of my comprehensive exams during the first two years of my studies. Their helpful advice and feedback helped me to design and run my first experimental syntax study on L1 and L2 processing.

Thirdly, I also wish to thank faculty members and staff in the Department of Linguistics and Languages, LMBLab, and ARiEAL at McMaster University. I learned the principles of Neurolinguistics and cognitive science of language from John Connolly and Lili Service. John and Lili's unique methods of teaching and their deep and recent knowledge in the field were so inspiring for me. I would miss

John's friendly manner, smiles, and chats. Thanks to Victor Kuperman's research methods and data analysis course, I was able to gain a deep understanding of designing empirical studies and applying mixed effects models in analyzing the data of my studies. He also took me to the world of visual language processing and the application of eye-tracking technique through his visual language processing course. I had the opportunity to audit his course on corpus linguistics, and familiarized myself with the principles of the field. I would like to thank Ivona Kucerova for teaching me new concepts and theories of syntax through her advanced graduate syntax course. Although the course was demanding and challenging, it prepared and helped me to design my first experimental syntax study. Thanks to Vikki Cecchetto's second language acquisition course, I was able to gain deep knowledge of first and second language acquisition and processing. This course was useful in designing and running four empirical studies on bilingual processing and performance. I had the chance to extend my knowledge of phonology and morphology by taking Anna Moro's graduate phonology and morphology course. Anna's unique way of teaching advanced phonology and morphology concepts and theories prepared me to work more efficiently as a teaching assistant for introductory linguistics, phonetics, and phonology courses. I had the honor to work as a phonology and phonetics teaching assistant for John Colarusso for four semesters. Besides his strong background and astounding knowledge in phonological analysis and phonetics, John is such a well-mannered and supportive professor to work with. I would like to thank Magda Stroinska for all her kindness and support as the head of the department. She never missed greeting me with a smile and asking if my family members were doing well. I appreciate Catherine Anderson for familiarizing me, as her teaching assistant, with new teaching methods through her blended teaching methods in her lectures. I was able to apply some of them in my tutorials and classes later. I would like to thank Daniel Pape for his support and help during my studies. Daniel was so helpful and supportive when I was digitizing my stimuli. Also, I appreciate Aki Kyröläinen's help and support on data analysis. His suggestions on applying mixed effects logistic regression models in analyzing the data enhanced my work a lot. Thank you, Wendy D'Angelo and Nikolai Penner, for all your support and encouragement. I would like to express my gratitude to Nanci Cole of the Department of Linguistics and Languages. Nanci was so helpful and professional whenever I needed a document or wanted to submit a request or application. You would never meet her without a smile, and you would never leave her office without the job done. I appreciate all Bianca James's help and efforts as the department's coordinator. Renee Wang is appreciated for her collaboration in translating English instructions to Mandarin for our Chinese-English bilingual study.

My fellow graduate friends had a great role in helping me to accomplish my goals. I commenced studying at McMaster without knowing anyone, but found a lot of nice friends with wonderful backgrounds. I would like to thank Ala Al-kajela, Jitka Bartosova, Laura Beaudin, Rober Bushera, Cassandra Chapman, Dian Doran, Kaitlin Faulkaukas, Angela Harrison, Gina Henrry, Amanda Ho, Connie Imbault, Olena Kit, Kenny Luk, Richard Mah, Ana Nunes, Gaisha Oralova, Fareeha Rana, Kyle Ruiter, Mackenzie Salt, Sheryl Sawyer, Bryor Snefiella, Daniel Smitchke, Heather Stephens, Narcisse Torshizi, Zoe Waelchi, and Chelsea Whitwell (in alphabetical order) for all their support and kindness. Special thanks to Daniel Schmidtke for all his support during my studies. Daniel was so encouraging and helpful when I was initially analyzing the data of my first experiment for my comprehensive exams.

The members of the LMBLab and some undergraduate volunteers helped me to record English stimuli, score non-word repetition tasks, run and score the Chinese tasks. It was a great experience for me to mentor some undergraduate students, involve them in my experiments, and have them co-present the findings of my studies. I would like to thank Victoria Stevenson, Courtney Tidd, Semona Basin, Jennifer Thammasouk, Runzie Li, Minhao Mao, Serena Ho, Allison Chen, Twinkle Chu Chen, and Yinan Zhang for all their cooperation. Also, I would like to thank Hamideh Nasseh, Afsaneh Sasani, and Nan Ma for reading Persian and Chinese stimuli for recording the stimuli for our experiments. I would like to thank all bilingual and monolingual participants who took part in our studies.

I would like to thank Michael Wilson of the McMaster University Research Ethics Board for his cooperation to review and approve our studies. I am grateful to the donors of the Elder Family Graduate Award for their merit-based graduate grants.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

AoA Age of acquisition

BI/BILING Bilingual

CH Chinese

ENG English

L1 First language

L2 Second language

L3 Third language

M Mean

MoA Manner of acquisition

NES Native English Speaker

OSpan Operation span

PER Persian

PLDB Persian Linguistic Database

PSTM Phonological Short-Term Memory

SD Standard deviation

SOV Subject object verb

SVO Subject verb object

WM Working Memory

WMC Working Memory Capacity

1 Introduction

The impact of language dominance and proficiency on linguistic and information processing and performance in bilinguals have been the subject of investigation in recent decades. A battery of early studies by Dornic (1980) showed that both decoding (comprehension) and encoding (production) are affected by language dominance, i.e. whether the task is performed in the participant's strongest (dominant) or a weaker (non-dominant) language. Participants were consistently slower in linguistic tasks such as language comprehension and production tasks, and also in other cognitive tasks such as digit detection or naming tasks, especially in complex tasks, in the second language (L2). Dornic believed that when tasks are done in the non-dominant language, this adds to the short-term memory load, and causes deterioration of performance.

The effect of language dominance, and its associated computational and cognitive loads on processing and performance, is further supported by studies exploring cross-linguistic language processing (Roberts, 2012), thinking abilities in the first language (L1) and second language (Manalo & Sheppard, 2016; Takano & Noda, 1993), arithmetic processing in L1 and L2 (Marsh & Maki, 1976; McClain & Huang, 1982; Van Rinsveld, Dricot, Guillaume, Rossion, & Schiltz, 2017; Van Rinsveld, Schiltz, Brunner, Landerl, & Ugen, 2016; Wang, Lin, Kuhl, & Hirsch, 2007), and the cognitive consequences of using a foreign language in the workplace (Neeley, 2013; Volk, Köhler, & Pudelko, 2014). While the L1 processing and performance are almost automatic and free of language-based computational costs, information processing when performing a task in a non-native language puts a strain on the cognitive system and requires more available memory resources (Green, 1998; Linck, Osthus, Koeth, & Bunting, 2014; Meschyan & Hernandez, 2006; Perani & Abutalebi, 2005; Roberts, 2012).

Variation in individual differences in working memory (WM), a limited capacity system that is responsible for the temporary maintenance and simultaneous processing of information (Baddeley, 2003; Baddeley, 2017; Cowan, 2005; Engle, Carullo, & Collins, 1991), and phonological short-term memory (PSTM) can affect language processing and performance. Research shows that there is a relationship between individual differences in working memory capacity (WMC) and language processing and performance, with advantages for individuals with greater working memory spans having more efficient processing and superior performance in sentence comprehension, ambiguity resolution, pragmatic or lexical-semantic information integration, and more sensitivity to morphological or syntactic violations (Dai, 2015; Farmer, Fine, Misyak, & Christiansen, 2017; Hopp, 2014; Just, Carpenter, & Keller, 1996; Kim & Christianson, 2017; Peng et al., 2018; Sagarra, 2017). On the other hand, phonological short-term memory has been associated with L2 vocabulary and grammar learning and oral fluency (Ellis, 1996; Martin & Ellis, 2012; Service, 1992, 2012). L2 learners who have better PSTM abilities demonstrate superior skills in L2 acquisition, learning, and performance.

Variables in bilinguals' language background can have a role in L2 processing and performance as well. Research has found that the onset of the age of L2 acquisition (AoA) (Archila-Suerte, Zevin, & Hernandez, 2015; DeKeyser, 2013, 2017; Delcenserie & Genesee, 2017; Jaroslawska, Gathercole, & Holmes, 2018; Roncaglia-Denissen & Kotz, 2016; Sabourin, Brien, & Burkholder, 2014; Sakai et al., 2009), the degree of L2 functional proficiency (Bel, Sagarra, Comínguez, & García-Alcaraz, 2016; Keating, 2017; Liang & Chen, 2014; Rossi, Diaz, Kroll, & Dussias, 2017), and L2 exposure and use (Christiansen & Chater, 2016; Deng, Dunlap, & Chen, 2017; Frenck-Mestre, 2002; Kroll, Dussias, Bice, & Perrotti, 2015; Pliatsikas & Marinis, 2013) can affect linguistic and information processing and performing real-life tasks in L2. Thus, L2 learners who acquire an additional language earlier in life, have higher L2 skills, are exposed to more L2 input and/or use L2 more frequently in daily activities or for professional responsibilities are more likely to demonstrate native-like processing and performance.

However, it is not yet fully clear how factors such as language dominance, linguistic complexity, individual differences in memory measures, and variation in linguistic history can affect following spoken instructions in bilinguals' dominance and non-dominance languages.

The present study

The main goal of the thesis presented here is to investigate the role of the language of the task, linguistic complexity, an individual's working or phonological memory capacity, and factors in an L2 speaker's linguistic background in encoding, decoding and performing verbal instructions in L1 and L2. I was particularly interested in the comprehension and performance of imperative sentences presented as sequences of oral instructions, which are sentence-level materials by nature, with readers or listeners required to comprehend the sentences and execute them. In addition to linguistic and information processing, the instruction-following task requires participants to actively hold the information in memory available for later retrieval and performance. This process is cognitively costly which may manifest in improper mental representations, forgetting the order of events, inaccurate recall, failure in performance, longer performance times, and higher error rates.

Previous research on instruction-following ability has mainly been done on native speakers of English, mostly on children. A few studies have been done on typical and atypical adults, (Allen & Waterman, 2015; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008; Jaroslawska, Gathercole, Allen, & Holmes, 2016; Jaroslawska et al., 2018; Waterman et al., 2017; Yang, Gathercole, & Allen, 2014; Yang, Allen, Holmes, & Chan, 2017; Yang, Allen, & Gathercole, 2016). The findings of these studies suggest that participants have more difficulties in verbally recalling the sequences of written or oral instructions than physically performing them. Younger children and subjects with lower working memory abilities or working memory impairment have been reported to perform more poorly in instruction recall.

This thesis contributes to knowledge about linguistic and information processing in bilinguals. To the best of my knowledge, no within-subjects studies of bilinguals have explored the role of language dominance, individual differences in memory, language history, and task complexity in following spoken instructions, so far. I investigated (i) whether the language of presentation (L1 vs. L2) affects recall accuracy in following sequences of simple and linguistically complex instructions. Furthermore, the results of previous studies on the effect of the language of testing on L1 and L2 WMC are mixed. Whereas some studies suggest that WMC is language independent (Harrington & Sawyer, 1992; Osaka & Osaka, 1992; Osaka, Osaka, & Groner, 1993), other studies have shown that L2 WM span is affected by language dominance and proficiency, with a lower WM span in the less skilled language users (Reichle, Tremblay, & Coughlin, 2013; Service & Tujulin, 2002; van den Noort, Bosch, & Hugdahl, 2006). In the current study, I examined the hypothesis that WMC measured by presenting a dual-task combining memory and another cognitive task (i.e. complex span task) is affected by the language of testing. To this aim, I created the testing materials in participants' first language (Persian in Experiments 1 & 3, Chinese in Experiment 2) and in English as the L2. To avoid learning and language confounds, I tested bilinguals in two separate sessions with at least a one-week interval. (ii) In Experiment 3, I manipulated the linguistic complexity of temporal adverbials. I embedded a semantically complex two-clause sentence in the sequences of four instructions to see to what extent the task difficulty impacts the ability to follow oral instructions in L1 and L2. I further investigated (iii) the possible relationship between WMC, PSTM, and following oral instructions. Lastly, (iv) I examined if the variables in bilinguals' language history, e.g., AoA, proficiency, and language use, can predict the instruction-following ability in bilinguals.

The hypotheses are that (i) language dominance influences the instruction-following ability and WMC in bilinguals, with an advantage for tasks presented in L1. (ii) Task complexity overloads working memory, affecting the recall accuracy and performance. (iii) Individual differences in language-specific WMC is correlated with the corresponding instruction-following skills, with an advantage in remembering and following instructions for subjects with more working memory resources. (iv) Language background variables, such as the onset of the age of acquisition and proficiency influence L2 comprehension and performance.

The remaining sections of this dissertation are organized as follows. In Chapter 2, I report the findings of the first experimental study, which examined the effect of language of the task (dominant vs. non-dominant), along with individual differences variables, on following spoken instructions in a group of skilled Persian L1 bilinguals and a group of native speakers of English as the controls. The empirical study in Chapter 3 investigated the predictive power of language history, WMC, and PSTM in instruction-following ability in a group of less-skilled Chinese L1 bilinguals and a group of native speakers of English as the controls. The experiment in Chapter 4 examined the impact of the linguistic complexity of sequential temporal order expressions, along with variation in individual differences in WMC, PSTM, and language background on following sequences of spoken instructions in a group of proficient Persian-English bilinguals and a group of native speakers of English. Chapter 5 includes a summary of the findings of the three studies and ends with a brief conclusion.

2 Experiment One

Following multi-step spoken instructions, the effect of language dominance

2.1 Introduction

The first experiment tests the hypothesis that being able to follow a series of verbal instructions is affected by language skill. Following instructions and directions is part of our real-life daily activities and functioning, taking place when responding to communication in the workplace, education, and training. For instance, technicians or workers need to follow instructions on how to operate equipment, patients need to remember the schedule and doses of medications they are taking, students need to follow their teachers' or professors' instructions in the learning process, and passengers need to know how to follow instructions in an emergency. However, following multi-step instructions is not as simple as it seems on the surface. Instructions are discourse-level utterances, so their execution primarily involves the application of the mechanisms of language processing, i.e., constructing meaningful mental representations from sentences unfolding in real time. However, in addition to language processing and constructing coherent mental representations, following instructions relies on additional cognitive resources to retain, interpret and execute the instructions. The listeners or readers have to actively maintain the information in memory, sometimes inhibit the immediate execution of current instructions, and based on incoming information, update and integrate the information with previous instructions. Finally, they have to recall the instructions when needed to be able to perform them in an orderly manner. Thus, encoding and temporarily maintaining the critical information in memory, inhibiting the execution of an instruction that is irrelevant or no longer relevant, recalling the information, and carrying out the actions, draw on a variety of higher-order cognitive control functions. These include monitoring, sustained attention, planning, inhibition, and task switching. The cognitive control functions enable the listeners or readers to remember the actions, attend to the current task, plan for sequential actions, to inhibit immediate execution of the current action when needed, and to switch flexibly between different steps to reach the goal (Yang, Allen, Holmes, & Chan, 2017).

Variables associated with following instructions have been reported in typical and atypical adults and children (Allen & Waterman, 2015; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008; Jaroslawska, Gathercole, Allen, & Holmes, 2016; Jaroslawska, Gathercole, & Holmes, 2018; Koriat, Ben-Zur, & Nussbaum, 1990; Lui et al., 2018; Waterman et al., 2017; Yang, Gathercole, & Allen, 2014; Yang, Allen, & Gathercole, 2016; Yang et al., 2017). In most of these studies, participants read or listened to sequences of instructions, and, either physically performed, or, verbally recalled them in the order they had been presented. The results of these studies revealed that recall accuracy was significantly higher when participants acted out the instructions, either with real physical objects or in a computer simulation version than when they just verbally repeated them. In Yang et al.'s (2016) study, which examined the immediate recall of sequences of spoken instructions in adult native speakers of English, the recall accuracy was affected by the type of recall task. In all three experiments, regardless of the experimental condition, participants had more accurate responses when they had to perform the sequences of instructions at recall rather than to verbally repeat them. For example, in their first experiment, the average of recalled enactment responses was 3.86 out of 5, versus 2.95 for verbal recall responses. The same performance advantage over verbal recall was observed in the Allen & Waterman

(2015) study. Twenty-eight adult native speakers of English heard sequences of instructions, and either performed or verbally recalled, the instructions in the correct sequences. The results showed that the effect of response type was significant, with enacted recall being more accurate than verbal recall.

Research has found that some individual difference factors, such as working memory capacity (WMC) and age, can affect the ability to follow written or oral instructions (Engle, Carullo, & Collins, 1991; Gathercole et al., 2008; Jaroslawska et al., 2016, 2018; Yang et al., 2014; Yang et al., 2016). Following instructions has been found to be correlated with tasks reflecting the efficiency of working memory (WM), a limited capacity system that is responsible for the temporary maintenance and simultaneous processing of information (Baddeley, 2003; Baddeley, 2017; Cowan, 2005; Engle et al., 1991). Information that must be kept available for ongoing task performance is thought to be kept in working memory. Central WM capacity is potentially limited by different types of factors (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). It has been suggested to be able to store limited amounts of information, 3-5 chunks (Cowan, 2001; Unsworth & Engle, 2007) or even fewer (Cowan, Saults, & Blume, 2014). The immediate and accurate recall is also available for a limited amount of time (less than one minute). These limitations, which impose restrictions on the processing and storage of information, are the origin of individual differences in WMC, and make some tasks more demanding for some individuals. Thus, it is predicted that participants with greater WM resources will outperform individuals with poor or impaired memory in tasks that impose a WM load. WMC, as a source of cognitive individual differences, has been found to be associated with performance in a range of complex cognitive tasks, such as language comprehension (Daneman & Carpenter, 1980), reasoning (Conway et al., 2005), simultaneous language interpreting (Macnamara & Conway, 2016), second language comprehension and production (see Linck, Osthus, Koeth, & Bunting, 2014 for a meta-analysis), and problem-solving (Price, Catrambone, & Engle, 2007).

Age is another critical factor that is correlated with both WMC and ability to follow instructions. Primary studies on children revealed that WM supports the retention and recall of instructions. In an early study, Engle et al. (1991) examined the role of individual differences in WMC in comprehension and following directions in first, third, and sixth graders. Memory ability was measured by a simple word span and Daneman and Carpenter’s (1980) reading span. The results showed that WM measures were correlated with the reading comprehension test and the ability to follow oral directions in the classroom. However, the ability to follow directions was also modulated by age. The complexity of the directions affected children's performance, with a disadvantage for younger and low-span participants. In another study, Gathercole et al. (2008) investigated following spoken instructions in laboratory analogs of classroom activities in 5- and 6-year-old children. Children listened to oral instructions involving the manipulation of a sequence of objects. They then either performed the instructions or repeated them. In a span-type procedure, the length of the instruction sequence was systematically increased up to the point at which the children could no longer perform the task. The results showed that performance was more than twice as accurate in the action than the verbal repetition condition. There was also a strong correlation between working memory ability and the accuracy with which children could carry out verbal instructions. It seems that some classroom activities, including following instructions, place a heavy demand on working memory. Thus, children with poor WMC may fail to meet the WM demands, which results in the forgetting of task-relevant information and, ultimately, task failure (Gathercole et al., 2008; Gathercole, Lamont, & Alloway, 2006).

In adults, evidence from following instructions in simple and dual task studies has revealed that working memory supports the retention and execution of both written and oral instructions (Engle et al., 1991; Gathercole et al., 2008; Jaroslawska et al., 2016, 2018; Yang et al., 2014; Yang et al., 2016). It seems that the ability to act out, or verbally repeat, sequences of instructions is correlated with simple and complex WM span measures, such as digit span and backward digit recall that tap either the storage or both storage and processing components of working memory. Some studies applied the dual-task (concurrent) paradigm to disrupt different components of working memory while participants were involved in instruction-following tasks. In three experiments, Yang et al. (2016) studied the involvement of WM in following spoken instructions in adult native speakers of English. In concurrent tasks, they applied a range of manipulations to disrupt different components of WM, for instance, the phonological loop (articulatory suppression task), central executive (backward counting by threes task), and visuospatial sketchpad (spatial tapping and blocked visual display tasks). Their findings revealed a reliable enactment benefit over the verbal recall of instructions. They found a close relationship between WM resources and the ability to follow instructions, and, further, that following instructions depended on multiple WM resources. However, memory for enactment was not affected by any of the secondary tasks. In three dual-task paradigm experiments, Jaroslawska et al. (2018) examined the immediate memory for sequences of spoken instructions (action-based vs. verbal recall). In addition to using secondary tasks (articulatory suppression and backward counting) during encoding and maintenance to tax the phonological loop and the central executive, they selectively disrupted the action advantage with concurrent motor suppression. The motor tasks involved producing a short repetitive sequence of fine motor gestures with the dominant hand (Experiment 1) and a more basic task that involved three gross motor gestures (Experiments 2 & 3). Consistent with previous findings, there was a recall advantage of enactment of instructions over verbal repetition. The verbal recall was affected by secondary tasks. However, the benefit of action-based recall was reduced following the production of basic gestures during encoding. Thus, the role of WM resources in instruction encoding and storage remains unclear.

The Present Study

So far, most studies have focused on following instructions in children's or adult participants' first language (L1). A few studies of working memory performance have investigated the role of storage and processing functions in the second language (L2) comprehension and processing by means of complex memory span tasks, mainly reading or listening sentence span. The results of such studies have been mixed. Some research found that there are no differences in WM capacity in L1 and L2, suggesting that WM capacity is language independent (Harrington &

Sawyer, 1992; Osaka & Osaka, 1992; Osaka et al., 1993). However, the results of other studies (Coughlin & Tremblay, 2013; Service, Simola, Metsänheimo, & Maury, 2002; van den Noort, Bosch, & Hugdahl, 2006) revealed that multi/bilinguals have larger WMC in L1 and their L2 or L3 WMC is determined by L2 or L3 proficiency. A study by Service et al. (2002) revealed that whereas highly proficient L2 users performed equally well as native speakers, less skilled bilinguals had lower reading span scores, indicating that lower proficiency in L2 consumes L2 learners' working memory resources. In another study, van den Noort et al. (2006) tested a group of L1 Dutch, L2 German, and L3 Norwegian multilinguals on simple and complex memory span tasks. The results revealed differences in performance in L1, L2, and L3. Participants had the largest functional WM capacity in L1 followed by L2 and then L3. WM resources depended on language proficiency. However, these studies involved artificial WM measures, and, it is not clear how the processing (executive control) and storage (capacity/span) functions of WM are affected when bilinguals carry out linguistic and non-linguistic tasks in daily activities in their non-dominant language.

Based on the assumption that functional WM capacity would be smaller in L2 than L1, the present study investigated whether bilinguals' dominant and non-dominant languages affect following sequences of spoken instructions. Measures were chosen to answer two questions: (1) whether there are differences in bilinguals' recall and performance of spoken instructions in L1 and L2, and, (2) whether individual differences in WMC or variables in bilinguals' language background have a detectable effect on the ability to follow oral instructions. To our knowledge, this is the first study of this type to examine the effects of the language of presentation, individual differences in WMC, and language background variables on following multi-step sequences of verbal instructions in bilinguals' native and non-native languages.

To answer the second question, we included individual differences measures, such as non-word repetition and complex span, to evaluate the relationship between WM and memory for instructions. Unlike simple memory measures like digit span, complex memory span measures, such as sentence span and operation span tasks, incorporate both processing and storage demands and can predict performance in higher-level cognitive tasks (Engle et al., 1991). In the original Daneman and Carpenter (1980) task, the processing component of the complex task was similar to the task to be predicted (reading comprehension). However, in Turner and Engle's (1989) operation span task, the processing component, i.e., the arithmetic computation, was different from the tasks being predicted. Yet, the individual differences in working memory capacity, measured by the operation span task, were found to be significantly correlated with higher-level cognitive functioning: reading comprehension. One significant motivation for using the operation span task in the present study was to see whether it would be correlated with memory for instructions irrespective of recall modality.

We employed a two-session (L1 vs. L2 within-subjects) design for this study. This enabled us to manipulate the language of testing in WM and instruction tasks (except the phonological memory measure that was presented in English only). Thus, any differences in participants' results were expected to reflect the consequences associated with language dominance (L1 vs. L2), variability in WMC (low vs. high), linguistic items, and participant variables (e.g., age of acquisition (AoA), proficiency, language use, etc.). We used modified versions of the instruction-following tasks of Yang et al. (2014). In their study, participants listened to sequences of instructions and either physically performed or verbally repeated the instructions in their serial order. We computerized the tasks to mimic real-world digitized activities. In addition to working with stationary objects, our participants performed some simulated office work, for example, copying or printing certain documents, and then putting them into specified folders. We hypothesized that (i) the language of presentation would influence bilinguals' ability to remember and follow oral instructions, giving rise to a disadvantage in tasks presented in the participants' non-native language, especially, for less skilled L2 learners. (ii) We expected positive correlations between performance in WM tasks as well as variables in bilinguals' language background, and the ability to remember and follow verbal instructions. In other words, working memory span and L2 variables were expected to predict bilinguals' immediate memory performance.

The participants were native speakers of Persian. Persian, a member of the Indo-European language family, is a Subject-Object-Verb (SOV) language with post-nominal adjective-modifiers, for instance, *poš ye qermez* (folder (the) red: the red folder). Thus, word order in Persian differs from that in English. As in English, imperative sentences such as instructions, directions, and commands are constructed with imperative verbs. Commands can be constructed either by adding the particle *be* to the bare root of the verb *(e.g., be gozar,* put*),* or without the particle in verb phrases containing light verbs like *kærdæn* (to do), (*e.g., print kon,* (the) printing do: print*).* Thus, a typical Persian command or instruction would be:

Xætkeš e	*abi*	*ra*	*daxel e*	*sæbæd e*	*gerd*	*be-gozar-id*
ruler	blue	OBJ-marker	into	basket	round	IMP PTCL-put-2 SG.

'Put the blue ruler into the round basket.'

2.1 Methods

2.1.1 Participants

Thirty-six Persian-English adult bilinguals (18 females and 18 males; age *M* = 27.75, *SD* = 6.79, range = 18 – 44 years) participated in the current study. Participants were recruited from undergraduate and graduate programs at McMaster University by posting the recruitment posters on Persian students' and Persian community Facebook pages and also displaying recruitment posters across the campus. They were paid 15 Canadian dollars per hour. Table 2.1 summarizes the English-Persian bilingual participants' language background information. Forty-nine dominant speakers of English (42 females and 7 males; age *M* = 20.47, *SD* = 4.50, range = 18 – 25 years; 79% born in Canada and 21% moved into the country before the age of 5) participated in this study as a control group. They were recruited via the SONA participant pool of the Department of Linguistics and Languages and received course credits for their participation. All participants had a normal or corrected-to-normal vision, normal hearing, and none reported a diagnosed colour blindness. The study was cleared by McMaster University Research Ethics Board.

Table 2.1 Summary of Persian-English bilinguals' language history.

Variable	M (SD)	Range	Variable	M (SD)	Range
Age of arrival	24.97(7.59)	7.84 − 43.75	AoA	10.40(2.87)	4.75 − 15.75
Length of stay	2.78(3.25)	0.08 − 16.16	Proficiency	4.47(0.70)	1 − 6
MoA:			Daily use		
Classroom	8.79(3.67)	3 −15	L1	0.56(0.19)	0.05 − 0.80
Naturalistic	3.26(3.70)	0.08 − 16	L2	0.43(0.19)	0.20 − 0.95
Mental use			Preference		
L1	5.72(1.09)	3.16 − 7	L1	2.91(1.26)	1 − 5.33
L2	3.52(1.03)	3.16 − 7	L2	2.75(1.47)	0 − 6.66

Proficiency (1: very low – 6: native-like), language use in all daily activities (0.00 – 1), mental language use (1: never – 7: always), and language preference (1: never – 7: always) were assessed based on self-rating reports by bilinguals. Manner of acquisition (MoA), formal classroom instruction or naturalistic exposure to L2 in an L2 environment, the age of L2 acquisition (AoA), the age of arrival in an L2 country, and the length of stay in an L2 environment were reported in years.

2.1.2 Memory stimuli and tasks

The stimuli for the enactment and verbal recall tasks were imperative sentences in the form of instructions. We were interested in instructions since they are sentence-level materials by nature, with readers or listeners required to comprehend the sentences and execute them — a cognitively demanding task. Experimental stimuli consisted of 240 unique items presented in 4 lists. Each list consisted of 12 sequences of 5 instructions (see Appendix A), yielding 60 individual items in each list (see Table 2.2). The order of items in a given sequence within a list was fixed.

However, we counterbalanced the order of conditions and also the assignment of stimuli lists for tasks and languages.

Table 2.2 Summary of the organisation of stimuli.

List 1 … List 4	
Sequence 1	Instruction 1
	Instruction 2
	Instruction 3
	Instruction 4
	Instruction 5
. . .	. . .
Sequence 12	Instruction 1
	Instruction 2
	Instruction 3
	Instruction 4
	Instruction 5

The pool of words used to create the sentences included 7 action verbs, 13 objects (nouns), 7 types of documents, 5 colours, and 2 physical shapes (round/square). Colour and shape features created competitors for similar objects. Each sequence consisted of five simple instructions. Two of simple instructions in each sequence were connected by the connectives "and then", involving two actions, e.g., *pick up the green ruler, and then put it into the square basket*. We avoided repeating the same verbs, objects, documents, colour or physical shape features in the same sequence of instructions. A typical example of a sequence of five instructions would be,

"*Move the yellow mouse onto the pad, and print the budget report, and then put it into the blue folder, and place the ruler into the round basket, and put the highlighter into the black tray".*

We controlled for the length of each sequence of instructions. The average number of words in each sequence was 34 with a range of 33 – 35 words.

The English sentences were recorded by a female native speaker of English, and were digitized on a computer using Audacity software with a sampling rate of 44100 Hz. In prosody, we emphasized the distinctive features between objects or containers, e.g., colours and shapes. Otherwise, the speaker was advised to follow natural prosody. The first four instructions were followed by 500 milliseconds (ms) of silence, and there was a 1000 ms silence after the last instruction. A beep sound followed the last silence after instruction five, signaling participants to commence recall.

For Persian tasks, the target sentences and experimental instructions were translated into Persian by a native speaker of Persian. The target sentences were recorded by a female native speaker of modern Persian (Tehrani dialect), and were digitized on a computer. The average number of words in each Persian sequence was 35 with a range of 34 – 37 words. We followed the same digital recording procedures as in the English task. However, each participant received two distinct lists for each language (enactment vs. verbal recall task) and the order of language of presentation and tasks were counterbalanced in sessions one and two.

We used Microsoft PowerPoint and SuperLab 4.5 (Cedrus) to design and create enactment and verbal recall tasks. Bilingual participants also filled out a customized language background questionnaire (see Appendix B) to collect demographic information, as well as information about their L2 (English) functional proficiency, L2 acquisition history, language use, and language preferences.

2.1.3 Individual differences measures

In order to explore the involvement of working memory capacity and language variables in instruction memory, a number of additional tasks were included. We tested participants' phonological memory using the non-word repetition task (Archibald & Gathercole, 2006), in which participants heard and said aloud 40 English-based non-words, ranging from 2 to 5 syllables in length (see Appendix C). The pseudo-words were digitally recorded onto a computer by a female native speaker of Canadian English, using Audacity software with a sampling rate of 44100 Hz. This task was scored by counting the number of correctly repeated syllables of the non-words.

We also used a modified version of the Operation Span Task (OSpan) (Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005) to measure participants' working memory capacity. The OSpan task, like the Daneman and Carpenter (1980) sentence span task, is a complex memory span measure that taps both processing and storage components of working memory. In this task, participants read aloud and computed sets of arithmetic operations and memorized a word for later recall in connection with each operation. In the most common OSpan task versions, the memory items are letters. However, as letters have different names in different languages, we opted out to use short words instead. For our version of the operation span task, we created a separate stimulus list for each language. The lists consisted of sets of 2 – 6 arithmetic operations with 2 – 6 words for memory (see Appendix D for English stimuli). This resulted in 60 individual equations. Participants read aloud and carried out sets of arithmetic operations and had to memorize words for memory. For instance, a sequence of length two could be:

(6 x 8) – 4 = 44, (Y/N), CHILD

(9/3) + 8 = 16, (Y/N), SNAKE

RECALL

The English words to remember were mono- and bi-syllabic nouns and were taken from the MRC Psycholinguistic Database Output (Science and Technology Facilities Council, 2016). The Persian words for memory were mono- and bi-syllabic nouns and were taken from the Persian Linguistic Database (PLDB) (Assi, n.d.). All words were checked for frequency, the number of syllables, letters, and phonemes. The order of items in a given sequence within a list was fixed; however, the sets within a sequence were counterbalanced so that participants could not predict the length of the upcoming sets. We counterbalanced the order of the assignment of OSpan lists for L1 and L2. We used SuperLab 4.5 software for programming and running the non-word repetition and OSpan tasks. Bilinguals were presented the OSpan tasks in their L1 and L2, with a one-week interval.

2.1.4 Design and procedure

The instruction-following experiment included two within-subjects variables: the language of instruction presentation (L1 Persian vs. L2 English) and the recall task (enactment vs. verbal recall). The order of items in a given sequence within a list was fixed. However, we counterbalanced the order of conditions and also the assignment of stimulus lists for tasks and languages. Thus, English monolingual participants were assigned to two of the four stimulus lists while bilinguals were assigned to two of the four stimulus lists in each language.

The participants were tested individually in the Language, Memory, and Brain Lab (LMBLab) at McMaster University. We tested the native speakers of English in a single session whereas bilinguals were assigned to L1 or L2 tasks in two separate sessions with a one-week interval, one in L1 and one in L2, with the language order counterbalanced between participants. In all tasks, the experimenter used the language of testing to greet, give experimental instructions and/or talk with participants. The L1 sessions were facilitated by the experimenter who was a native speaker of modern Persian. Thus, all the communications and instructions were in

Persian when the participants were attending an L1 session. To eliminate the possible confounding of effects of the language of presentation and learning, we counterbalanced the order of the tasks (enactment vs. verbal recall) and the language of testing (English vs. Persian). Upon arrival, the participants signed an informed consent form. They were naïve to the intent of the study and were told that they would listen to some sentences in the form of instructions in L1 or L2 for comprehension and would then either be asked to enact them in order or verbally repeat them. The participants were seated in front of a 21.5" Macintosh computer displaying pictures of a number of office-related objects on the screen (see Appendix E). They did a naming task to assure their familiarity with all objects, documents, colours, and shapes that would be included in the instructions before commencing the experiment. We also instructed the participants how to carry out the actions, e.g., “signing”, “copying”, or “printing” documents, and “picking up” objects and then putting them onto/into specified containers or positions. Then, participants were allocated to the enactment or verbal recall task in L1 or L2. Before beginning the main tasks, participants completed a practice list, where they acted out or verbally repeated two sequences of five instructions and had time to ask questions or request further clarifications.

We probed memory for instructions by either the enactment or verbal recall task. In the enactment task, participants saw pictures of objects, documents, and devices on each PowerPoint slide. They then double-clicked on the sound icon on the top left of each slide to listen to each sequence of instructions via headphones, trying to comprehend the instructions. They were instructed to memorize the instructions, their serial positions in the sequence, and other details, while instructions were being played. In the enactment task, the importance of performing the actions in correct sequence was emphasized. Upon hearing the beep sound after the last instruction in each sequence, participants had to use the computer mouse to immediately start executing the instructed actions on the computer screen, for example, dragging a picture of a document onto a picture of a printer or moving the picture of an object onto the picture of a container. Participants' enactments were screen-recorded by the Quick Time Player software for data collection, scoring, and analysis.

In the verbal recall task, after pressing a key on the keyboard, participants would see the pictures of the objects, documents, containers, and devices of each sequence on the computer screen. At the same time, each sequence of auditory instruction related to the objects would begin playing on the headphones. The participants were advised to listen carefully as the audio files would be played only once. Upon hearing the beep sound after the last instruction, participants recalled the instructions by saying them aloud in their serial order in the given sequence. The importance of repeating the instructions in correct order was emphasized. The pictures remained on the screen until participants pressed a key to proceed to the next sequence. Their voices were recorded by Audacity software for scoring and analysis.

The OSpan task followed the enactment or verbal recall task and was presented in the same language as the rest of the L1 or L2 session. Before running the experimental lists, participants did a practice list. The participants were asked to read aloud each arithmetic operation in the equations, perform the mental computation simultaneously and immediately press the specified Yes/No keys on the keyboard to decide if the provided answer was correct or incorrect. Upon pressing the Y/N key, they saw a word (a noun) in the upper case in red for 500 ms on the next screen. They were asked to also read it aloud as well. They then continued doing the sets until they had completed all items of a sequence and saw “RECALL” on the last screen. This signaled time to recall the words presented with each set of the equation by saying them in correct serial order. If they knew the sequential position of a word in a given sequence but were not able to recall it, they had been instructed to say “BLANK” for each missed word. The order of the sets of different lengths was counterbalanced so that participants might not be able to guess the length of each sequence. The experimenter sat at his desk and noted down the words the participants recalled and said aloud.

Reading times and arithmetic accuracy for each trial were collected by Superlab 4.5 software. Before running the experimental list, participants did a practice task consisting of two sequences. The experimental instructions, the numbers, and the words were in either L1 or L2, depending on the language of the testing session. Bilinguals were asked to avoid doing the arithmetic computation in L1 if they were attending the L2 session.

The non-word repetition task was presented in the English session. After a practice list, participants listened to individual non-words and repeated them aloud. The importance of careful listening and accurate repetition were emphasized. The participants' voices were recorded by Audacity software for scoring and analysis.

2.1.5 Statistical considerations

Two participants from the native speakers of English group were excluded from data analysis because of unreliable or missing data. For one subject, the non-word repetition task had not been recorded for technical reasons and the other subject skipped most instructions in the enactment task.

The dependent variable was recall accuracy in enacting or verbally recalling the individual instruction. We modelled the recall accuracy as the probability of correct response for each instruction in a given sequence, where a sequence is defined as a set of five oral instructions. We used a sequence length of five instructions based on previous studies, similar to Yang et al., (2016) study. However, contrary to most similar studies that have relied on averaging the performance of each subject in each task by employing ANOVA models for data analysis, we used the more powerful generalized linear mixed effects models with multiple variables and covariates. Given that the recall accuracy for individual instructions was bound by 0 and 1, violating the assumptions of linear regression models and ANOVA (see Jaeger, (2008) for discussion), we used the generalized linear mixed effects regression models (Baayen, 2008; Baayen, Davidson, & Bates, 2008) as implemented in the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) and R Core Team, (2017) with a binomial distribution to perform a linear mixed effects analysis of the relationship between independent variables and recall accuracy in following sequences of five spoken instructions. This method of analysis allows to explore the effects of multiple factors and covariates while separately accounting for any variance contributed by participants and items (in this case, each individual instruction). As fixed effects in the models, we had the language of presentation (Persian or English), group (monolingual or bilingual), the type of task (enactment or verbal recall) with scores on the complex working memory span, scores on phonological memory, and variables in bilinguals' language background as continuous covariates. In addition, we used the position of each instruction in a given sequence (1-5) and trial (running number of an item in a list, 1-60) as control variables. As random effects, we had intercepts for participants and items. Initially, we fitted each model with a maximal fixed effects structure and then removed the factors that did not significantly improve model's performance. Where a model comparison was done, p-values were obtained by the likelihood ratio test of the full model including the effect in question against the model without the effect in question. Separate analyses were run to compare performance on English instructions between groups (monolinguals and bilinguals) and to compare within-group performance in L1 and L2 in the bilingual group.

The enactment of an individual instruction was scored correct if participants carried out the correct action on the said object or document in its correct sequential order position among the set of five instructions. Each individual instruction in the verbal recall task was scored correct if the participants correctly repeated the instruction in its correct sequential position and with all details. Thus, an incorrect or missed instruction received zero points. In the enactment task, any errors in the sequential order of instructions, performing the wrong action verbs, moving the wrong objects or documents to the wrong containers or positions, or mixing up colour or shape features were the odds for getting an instruction correct and in the correct position out of five. In the verbal recall task, any errors in the sequential order of instructions, recalling the wrong action verbs, objects or documents, containers or positions, or colour or shape features were the odds for getting an instruction correct and in the correct position out of five. However, as some actions verbs such as “place”, “put” or “move” required similar actions in the enactment task, the response was acceptable if participants recalled and said a synonym or similar action verb in the verbal recall task. Also, participants were not penalized if they forgot to recall the function words such as articles or prepositions in the verbal recall task. The following is an example of a Persian-English participant’s responses in the English verbal recall task.

Sequence 1, List 1, English Stimuli

a. place the red stapler in the box

b. put the highlighter into the blue tray

c & d. sign the tax form, and then put it into the yellow folder

e. and put the black pen into the basket

Participant's recall:

a. put the stapler onto red stapler onto the **basket** (*wrong object error*)

b. put the **marker** onto the blue **box** (*wrong object and container errors*)

c & d. <u>sign the tax form</u>, and after that put it, sign the tax form by the blue pen and after that put it onto **the folder** (*folder color is missing, yellow or green?*)

e. No response

Only the recall for the third instruction (c) is acceptable.

The data collected from English and Persian operation span tasks were scored based on the number of memorized words recalled correctly in serial order within each set of equation-word pairs. The number of memory words correctly recalled in their serial positions in each sequence was divided by the total number of words in that sequence (2 to 6). The scores were averaged for each participant and henceforth, the results of the operation span task will be reported as “working memory span (WM span)” or “working memory capacity (WMC)”.

The data of the English non-word repetition task were scored by two student members of the LMBLab. They were native speakers of English and majoring in Linguistics or Cognitive Science of Language program. Participants received partial credits based on the accurate repetition of the number of syllables in each repeated non-word. The scores were averaged for each participant, and henceforth, the results of the non-word repetition task will be reported as “phonological memory”.

2.2 Results

2.2.1 Enactment and verbal recall tasks

As descriptive data in Table 2.3 shows, the language of testing seems to have influenced bilinguals’ performance in both enactment and verbal recall tasks, with an advantage for tasks done in L1. In addition, all participants consistently were more accurate in acting out instructions than verbally recalling them. The effect of the task appears present regardless of the language of presentation and group.

Table 2.3 Means and standard deviations of recall accuracy of acting out and verbally recalling sequences of five instructions.

		Enactment		**Verbal Recall**	
Group	**Language**	*Mean*	*SD*	*Mean*	*SD*
Bilingual (*N* = 36)	Persian	0.73	0.44	0.51	0.50
	English	0.62	0.49	0.49	0.50
Monolingual (*N* = 47)	English	0.68	0.47	0.45	0.50

For the bilingual group, we fitted L1 and L2 data into the generalized mixed effects models to examine the impact of the language of presentation (L1 vs. L2) on recall accuracy in following instructions. The results of the models revealed a significant main effect of the language of presentation on recall accuracy in bilinguals. The language of testing, as a within-subject factor, significantly influenced bilinguals' recall accuracy in acting out and verbally recalling sequences of five spoken instructions, with a disadvantage for instructions presented in L2 (see Table 2.4). There was also a significant main effect of recall task, with better performance for enactment than verbal recall. The position of the instructions in a sequence (1-5) significantly influenced the recall accuracy with instructions at the beginning of the sequence benefiting from the primacy effect. The running number of items in a list (1-60) did not significantly affect the recall accuracy.

Table 2.4 Summary of the final logistic regression model of bilinguals' (N = 36) recall accuracy in following L1 and L2 instructions, reported as the regression coefficient estimates, standard errors, z values, and p values. *Position* refers to position in sets of five. *Trial* refers to the running number (out of 60) of the trial as counted from the beginning of the experiment. The reference level is indicated for categorical variables.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.7361
Participant	Intercept	0.3745
	Language = *Persian*	0.3513
	Task = *Verbal recall*	0.4109

Part B: fixed effects

	Estimate	Std. Error	z value	p value
Intercept	0.5547	0.0919	6.037	<.001
Position	-0.1779	0.0380	-4.683	<.001
Trial	-0.0042	0.0031	-1.372	0.17
Language = *Persian*	0.6434	0.0988	6.514	<.001
Task = *verbal recall*	-0.6113	0.0888	-6.886	<.001
Language *Persian** task *verbal recall*	-0.5275	0.0973	-5.421	<.001

However, as Figure 2.1 and Table 2.4 show, the language of presentation interacted with the type of task with lower recall accuracy in the verbal recall task in Persian. Thus, the enactment recall in Persian was significantly better than English, but this difference was smaller for the verbal recall as the standard error ranges overlapped. A follow up paired comparison, using the two-sample t-test, revealed that bilinguals had a higher recall accuracy in acting out Persian instructions than English instructions, $t = 8.3178$, $p < .001$. However, the differences in recall accuracy in Persian and English verbal recall tasks were not as significant as the enactment tasks, $t = 1.7042$, $p = 0.0884$.

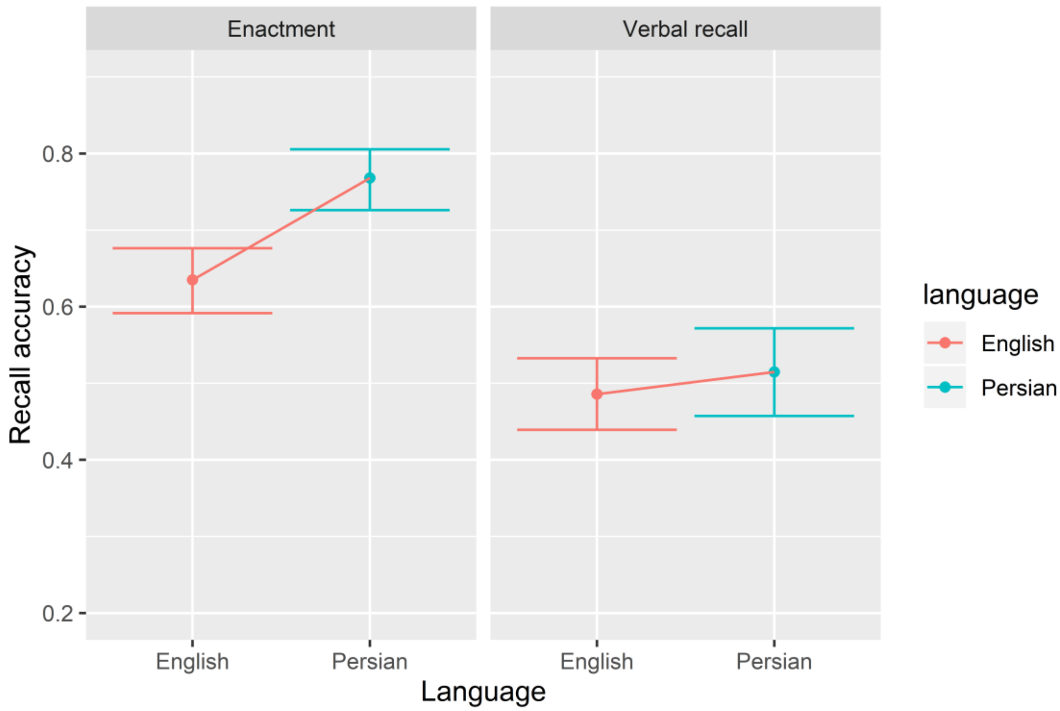

Figure 2.1 The interaction between language and the type of task in the bilingual group. Error bars show the standard error.

The type of task (enactment vs. verbal recall) affected the recall accuracy in L1 and L2 tasks. Bilinguals had lower recall accuracy in the verbal recall tasks in L1 (b = -1.1480, SE = 0.1075, z = -10.677, p < .001, model not shown) and L2 (b = -0.6132, SE = 0.1009, z = -6.073, p < .001, model not shown), suggesting the advantage of the enactment. As Figure 2.2 shows, the position of an individual instruction in a given sequence of five instructions affected the recall accuracy in L1 (b = -0.2239, SE = 0.0407, z = -5.497, p < .001, model not shown) and L2 (b =

-0.1366, SE = 0.0431, z = -3.169, p = 0.0015, model not shown), with recall accuracy being influenced by the primacy effect. However, the running number of each trial in a list of 60 instructions did not significantly influence the recall accuracy in acting out or verbally recalling the instructions in L1 and L2. This suggests that there was little effect of learning during the task.

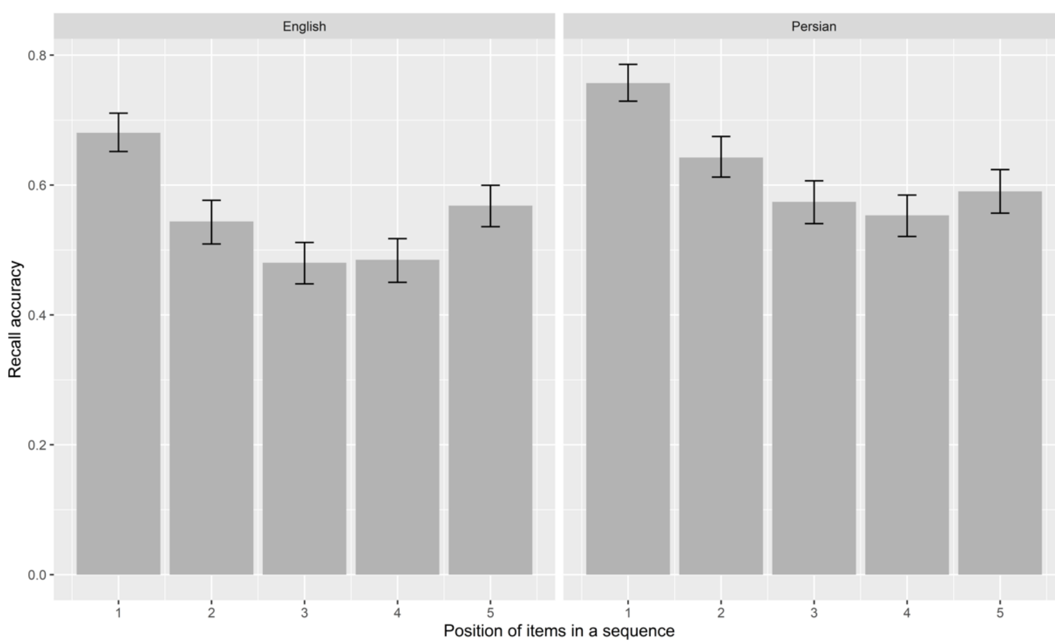

Figure 2.2 The effect of serial-order position of items in a given sequence of five instructions on recall accuracy in Persian L1 bilinguals. Error bars show the standard error.

In a separate model, we fitted the English monolinguals' data into the generalized mixed effects regression models to examine the relationship between the type of task and recall accuracy in following spoken instructions. The results of the models showed that there was a main effect of the type of task with significantly lower recall accuracy in the verbal recall task (b = -1.1623, SE = 0.1065, z = -10.918, $p < .001$, model not shown). The serial position of instructions in each sequence of five instructions also affected the recall accuracy, (b = -0.2657, SE = 0.039, z = -6.834, $p < .001$ model not shown), with higher recall accuracy in instructions in the beginning of the sequence. Again, the running number of trials in a list of 60 instructions did not significantly influence recall accuracy in either acting out or verbally recalling the instructions.

A separate analysis was conducted to compare the performance of bilinguals and monolinguals in only English instructions. As Table 2.5 shows, group as a between-subjects factor significantly affected the recall accuracy for sequences of English instructions. Bilinguals recalled fewer instructions than their monolingual peers. However, the disadvantage of bilingualism was limited to the enactment task because there was a significant interaction between the group and the type of task with the bilingual group having a higher recall accuracy in the verbal recall task than monolinguals.

Table 2.5 Summary of the final mixed effects logistic regression model of the effect of group, bilinguals (*N* = 36) and monolinguals (*N* = 47), on recall accuracy in English trials, reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to serial position of instruction and *Trial* to running number of trials from 1 to 60. The reference level is indicated for categorical variables.

Part A: random effects		
	Name	Std. Deviation
Item	Intercept	0.7308
Participant	Intercept	0.6340
	Task = *verbal recall*	0.5258

Part B: fixed effects	Estimate	Std. Error	*z* value	*p* value
Intercept	0.9193	0.1135	8.102	<.001
Position	-0.2053	0.0372	-5.519	<.001
Trial	-0.0047	0.0030	-1.554	0.1201
Task = *verbal recall*	-1.1494	0.0985	-11.666	<.001
Group = *bilingual*	-0.3577	0.1552	-2.305	0.0212
Task *verbal recall* * group *bilingual*	0.5297	0.1478	3.584	<.001

2.2.2 The influence of phonological WM: The Non-word Repetition Task

We investigated if phonological memory affected instruction recall accuracy. We inspected English-based nonword repetition scores and their relation to recall performance in bilinguals. First, we explored if mean phonological memory scores of bilingual and monolingual groups differed. The results of an independent-sample t-test revealed that monolinguals had higher phonological memory scores (*M* =

0.97, *SD* = 0.05) than Persian L1 bilinguals (*M* = 0.88, *SD* = 0.02), $t = -8.92$, $p < .001$. As Figure 2.3 shows, bilinguals were significantly less accurate in repeating English-based pseudo-words in the non-word repetition task than their native English monolingual peers.

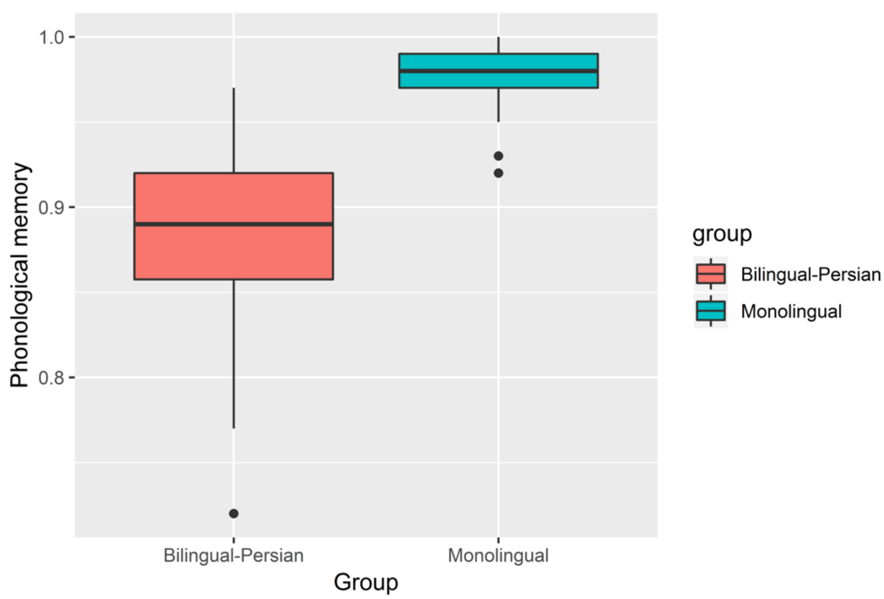

Figure 2.3 Comparison of phonological memory scores in Persian-English bilinguals and English monolinguals.

We further investigated if phonological working memory influenced recall accuracy in following spoken instructions in the English task in bilinguals. To this end, we fitted the data of the English enactment and verbal recall tasks and the phonological memory scores from the non-word repetition task into a series of generalized mixed effects regression models. The results of the final models predicting recall accuracy for English instructions revealed that the main effect of phonological memory on recall accuracy in the enactment and verbal repetition tasks in English trials was not significant in bilinguals (see Table 2.6). As Table 2.6 shows, the interaction between phonological memory and the type of task was not significant either. However, it seems that participants with higher phonological scores had better recall accuracy in the verbal recall task.

Table 2.6 Summary of the final logistic regression model of the relationship between instruction recall accuracy and phonological memory in bilinguals (*N* = 36), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number of trials over the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	0.7806		
Participant	Intercept	0.3756		
	Task = *verbal recall*	0.4370		
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	1.8218	1.3331	1.367	0.1717
Position	-0.1366	0.0431	-3.170	0.0015
Trial	-0.0034	0.0035	-0.971	0.3315
Phonological memory	-1.4316	1.5032	-0.952	0.3409
Task = *verbal recall*	-2.5333	1.6726	-1.515	0.1299
Phonological memory * task *verbal recall*	2.1712	1.8874	1.150	0.2500

2.2.3 The effect of working memory capacity: The Operation Span Task

We examined the hypothesis that working memory capacity correlates with recall accuracy in following oral instructions. Thus, individuals with larger WM span scores would have higher recall accuracy in enacting and verbally repeating sequences of instructions. Also, we further investigated if the language of the complex span task, operation span, influences WM span, assessed separately in L1 and L2. To this end, we measured working memory capacity in bilinguals' dominant and non-dominant languages.

As descriptive data in Table 2.7 suggests, the language of the task affected bilinguals' WMC scores in the complex span task. Bilinguals had better WM span scores in L1 complex span task than L2 complex span task. The average arithmetic accuracy in both bilingual (L1 and L2 OSpan tasks) and English monolingual groups was well above chance, suggesting that participants were actively engaged in the processing aspect of the complex span task, i.e., computing the output of the arithmetic operations.

Table 2.7 Summary of WM span in bilinguals (N = 36) and monolinguals (N = 47), reported as recall accuracy scores: proportions averaged across items, arithmetic accuracy, and mean RT of reading out equations and making verification decisions in milliseconds.

Group	Language	WM span (SD)	Arithmetic accuracy (range)	Mean RT (ms)
Bilinguals	Persian	0.73 (0.29)	0.94 (0.77-1)	7857
	English	0.67 (0.32)	0.94 (0.63-1)	9351
Monolinguals	English	0.66 (0.32)	0.89 (0.55-1)	7103

To find out any possible differences between L1 and L2 WM measures in bilingual participants, we conducted an independent-sample t-test. The results showed that the language of presentation influenced WM span scores, $t_{(69)} = 2.74$, $p = 0.0078$. Persian L1 bilinguals recalled more words for memory in the correct serial order in L1 trials than L2 trials (see Figure 2.4). Despite the differences in the scores of L1 and L2 complex span tasks, L1 and L2 WM spans were correlated, $r = .22$ [95% CI: -0.12 – 0.51].

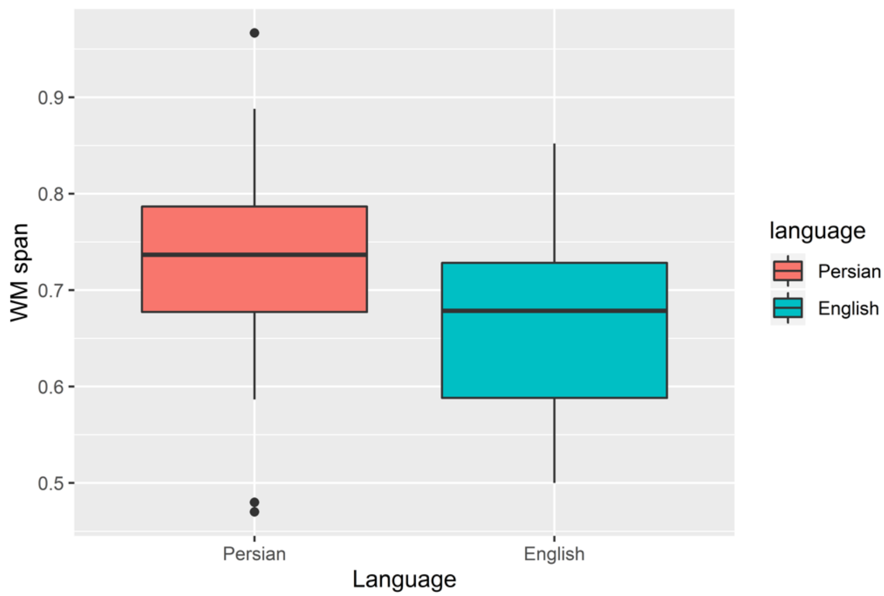

Figure 2.4 The effect of the language of presentation (L1 vs. L2) on WM span in Persian-English bilinguals.

However, in a model of English trials only, comparing monolingual English native speakers and Persian L1 bilingual speakers, the results of an independent-sample t-test revealed that there were no significant differences between the means of WM span scores in bilingual and monolingual groups, $t_{(81)} = -0.33$, $p = 0.7436$.

We further tested the hypothesis that there is a relationship between individual differences in WM span and recall accuracy in following verbal instructions. Thus, individuals with a greater WM span would have higher recall accuracy. We fitted the recall data of L1 and L2 enactment and verbal recall tasks and the scores of L1 working memory span into a series of generalized mixed effects regression models, with participants and items as the random effects and WM span as a fixed effect. The results of the final models showed that WM span affected instruction recall with an advantage for participants with a higher WM capacity (see Table 2.8 and Figure 2.5). Thus, bilinguals who had larger available working memory resources were able to retain, retrieve, and follow more sequential instructions. The interaction between WM span and the type of task was not significant, $b = 0.4470$, SE = 0.8333, z = 0.536, $p = 0.5916$, suggesting that recall accuracy in both enactment and verbal recall tasks were correlated with the WM scores. The comparison of models, using the likelihood ratio test, indicated that the interaction model was not necessarily a better fit than the simple model, X^2 (1) = 0.2848, $p = 0.5936$. Therefore, we are reporting the results of the simple model (see Table 2.8).

Table 2.8 Summary of the final logistic regression model of the relationship between instruction recall accuracy and L1 WM span in bilinguals (*N* = 36), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number of trials over the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name		Std. Deviation	
Item	Intercept		0.7331	
Participant	Intercept		0.3965	
	Language = *Persian*		0.3996	
	Task = *verbal recall*		0.3536	
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-0.3249	0.4975	-0.653	0.5137
Position	-0.1775	0.0379	-4.687	<.001
Trial	-0.0042	0.0031	-1.371	0.1705
Language = *Persian*	0.3606	0.0823	4.384	<.001
Task = *verbal recall*	-0.8612	0.0765	-11.257	<.001
WM span	1.3823	0.6716	2.058	0.0394

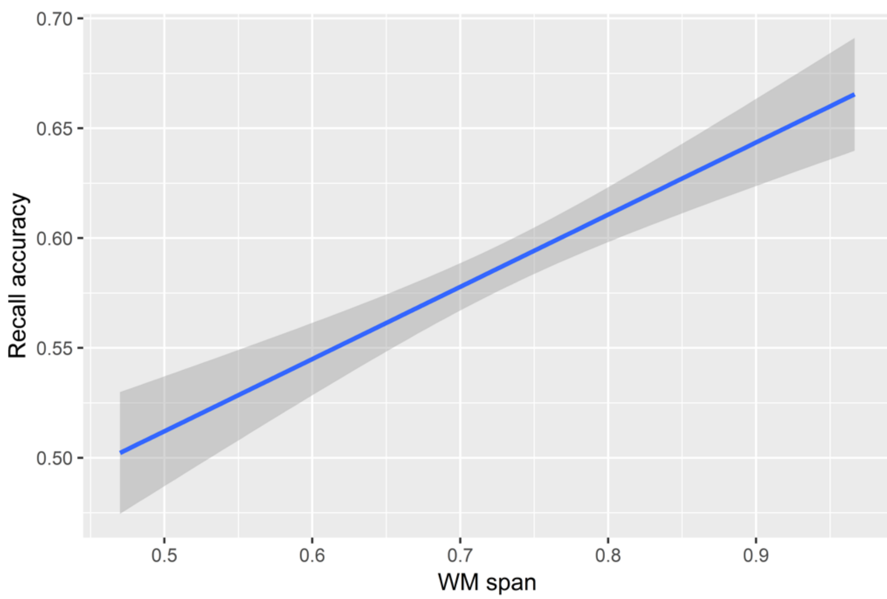

Figure 2.5 The relationship between L1 WM span and instruction recall accuracy in L1 and L2 tasks in Persian-English bilinguals.

We further investigated if there was an interaction between WM span and the language of instruction-following tasks in bilinguals. The results of the models showed that the interaction between language and WM span was significant, b = 2.4162, SE = 0.7084, z = 3.411, p <.001, model not shown. As Figure 2.6 shows, bilinguals with a higher WM span had better instruction recall accuracy in L1 than in L2.

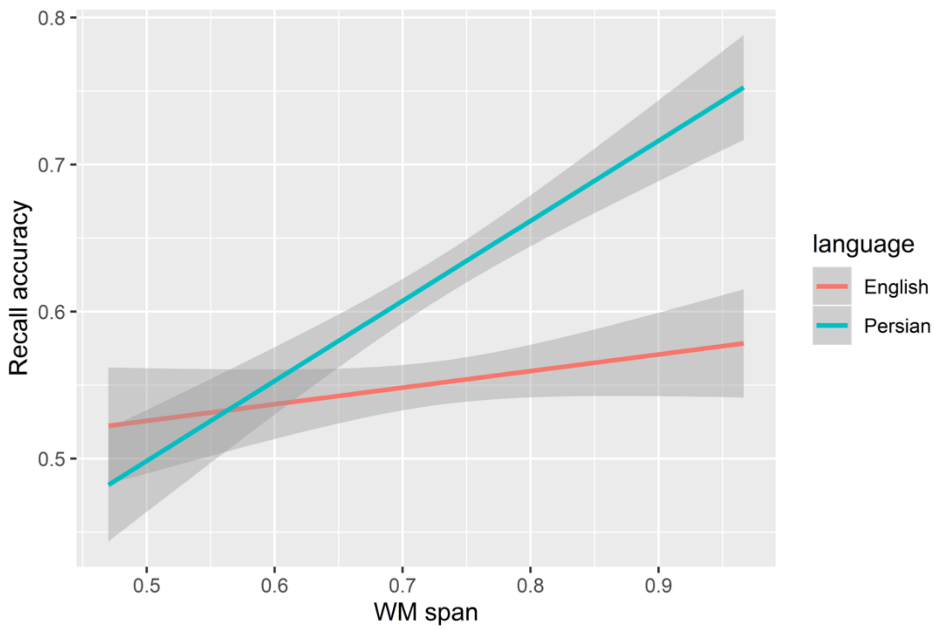

Figure 2.6 The interaction between L1 WM span and instruction recall accuracy in L1 and L2 tasks in Persian-English bilinguals.

In a separate model, we explored the relationship between WM span and instruction recall accuracy in English monolinguals. As Table 2.9 shows, there was the main effect of WM span on recall accuracy in both the enactment and verbal recall tasks in the English monolingual group. As Figure 2.7 shows, monolingual participants with greater working memory capacity had higher recall accuracy in following sequences of verbal instructions. The interaction between the scores of WM span and the type of task was non-significant (see Table 2.9).

Table 2.9 Summary of the final logistic regression model of the relationship between instruction recall accuracy and WM span in monolinguals (*N* = 47), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number of trials over the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	0.6995		
Participant	Intercept	0.7451		
	Task = *verbal recall*	0.5768		
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-0.5773	0.7113	-0.812	0.4170
Position	-0.2656	0.0389	-6.836	<.001
Trial	-0.0060	0.0032	-1.895	0.0580
Task = *verbal recall*	-0.4786	0.6389	-0.749	0.4539
WM span	2.2947	1.0656	2.153	0.0313
Task *verbal recall** WM span	-1.0381	0.9593	-1082.	0.2792

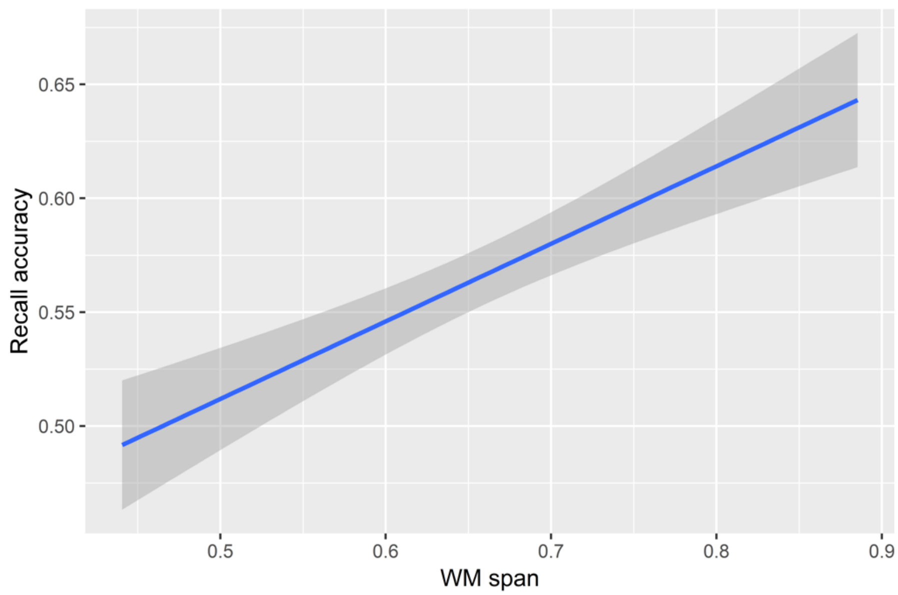

Figure 2.7 The relationship between WM span and recall accuracy in following spoken instructions in the English monolingual group.

In sum, the results of the models examining the relationship between individual differences in working memory and instruction recall accuracy supported the hypothesis that WM span was correlated with the ability to remember and follow oral instructions. In both English monolingual and bilingual groups, individuals with a greater WM span were able to both act out and verbally repeat a greater number of oral instructions than those with an inferior WM span. The fact that we found no significant interaction between WM span and the type of task indicates that the effect of WM span was present regardless of the type of the task.

2.2.4 L2 background variables

We further explored if there was a relationship between the variables in bilinguals' language history and the ability to remember and follow L2 verbal instructions. We fitted the data from L2 enactment and verbal recall tasks and language acquisition history into the generalized mixed effects regression models with variables in the L2 learners' language history as predictors. We had participants as a random effect and as fixed effects we had participants' current age, the level of education (undergraduate, Masters, and Ph.D.), gender, the age of arrival in Canada, the length of residence in Canada and/or an English speaking country (in years), the manner of acquisition of English as a second language (classroom instruction vs. natural exposure (in years)), the mean onset of L2 age of acquisition, the mean degree of functional proficiency in listening, speaking, reading, and writing skills (based on self-reported rating 1-6), L1 and L2 daily use (based on self-reported percentage), mental L1 and L2 use (based on self-reported rates 1-7), and L1 and L2 preference in academic and non-academic situations (based on self-reported rating 1-7). We found no significant effects of the fixed effects (age, education, gender, the age of arrival, length of residence, manner of L2 acquisition, L2 AoA, L2 proficiency, L1/L2 use, or language preference) on recall accuracy in the enactment and verbal repetition tasks. Furthermore, the elements in the participants' language history did not significantly interact with the type of task.

As we tested bilingual participants in two separate sessions, with one-week intervals, we conducted post hoc analyses to explore any possible effects of the session order (session 1 vs. session 2) on recall accuracy in the enactment and verbal recall tasks. The results of the mixed effects models, with participants and items as random effects, did not reveal any significant main effects of the session order on recall accuracy in L1 and L2 enactment tasks, $b = 0.0629$, SE = 0.1598, z = 0.394, $p = 0.6937$ (model not shown), and L1 and L2 recall tasks, $b = 0.0819$, SE = 0.1768, z = 0.464, $p = 0.643$ (model not shown). Furthermore, there were not any significant interactions between the language of testing and the session order in L1 and L2 enactment tasks, $b = 0.0091$, SE = 0.3119, z = 0.029, $p = 0.9768$ (model not shown), and the language of testing and the session order in L1 and L2 verbal recall tasks, $b = 0.5240$, SE = 0.3224, z = 1.624, $p = 0.104$ (model not shown).

2.3 Discussion

The present study is the first empirical investigation of the cognitive ability of bilinguals to follow spoken instructions in their two languages. More importantly, the potential consequences of performing and verbally repeating oral instructions in a non-dominant, but proficient, language were explored. We also examined any possible relationship between individual differences in working memory measures, variables in bilinguals' L2 language history, the type of task, and the ability to follow spoken instructions.

Although we tested a group of homogenous advanced learners of English, remembering sequences of instructions was affected by the language of presentation. Bilinguals' recall accuracy was lower, as reflected in the enactment and verbal repetition tasks, when the target stimuli were presented in the non-dominant language, English. As all the testing materials and procedures were similar in L1 and L2 sessions, the observed differences in bilinguals' performance appear to be the consequence of linguistic processing and performance in a non-dominant language. We found an L2 disadvantage in simulated office work. Thus, the results of our study support previous suggestions that linguistic and information processing in the context of executing professional responsibilities in a non-native language is cognitively demanding and puts a strain on the processing system (Roberts, 2012). As linguistic processing and constructing mental representations should be completed before performing the commands, the consequences of the language of task (dominant vs. non-dominant) would be reflected in lower recall accuracy in performing tasks. Earlier work by Takano and Noda (1993) observed that the processing load of English as a foreign language caused a temporary decline in the thinking ability of Japanese native speakers in a concurrent task. In their study, the L2 learners listened to factual questions and provided yes/no answers while doing an arithmetic calculation thinking task: the addition of two-digit numbers. The results of their study showed that participants had a higher correct response and lower error rates when the task was presented in the dominant language, Japanese. In the present study, the same effect was observed also in the complex span tasks, the operation span task. The language of testing significantly influenced bilinguals' working memory scores, with an advantage for recalling more words for memory in L1, Persian. The fact that bilinguals spent longer times on the L2 OSpan task suggests that doing a demanding task in a non-native language may deplete working memory resources, resulting in slower processing and difficulty in storing and retrieving information.

Our results on the effect of the language of presentation are further in line with studies exploring L2 arithmetic processing in unbalanced bilinguals (Marsh & Maki, 1976; McClain & Huang, 1982; Wang et al., 2007), suggesting that arithmetic calculation in a non-dominant language is slower and less accurate. In an fMRI study, Wang et al. (2007) investigated the neuro-mechanisms underlying mathematical processing in L1 and L2 in Chinese-English bilinguals. The results of their study revealed that calculation in L2 resulted in more errors and longer RTs than that in L1, indicating that doing the tasks in L2 was more difficult than in L1. In contrast to the native language, performing the calculation in a non-native language engaged additional and more extensive neural sites, particularly in the language-dominant left hemisphere, including the inferior frontal gyrus (Broca's area). They associated the activation of the left inferior frontal gyrus for the L2 with an effort to translate the problem from an unfamiliar language to a familiar language. Thus, the reliance of L2 calculation on language systems indicates that L2 input may have been translated into or mediated by L1 to perform the calculation.

The empirical design of our study, which integrated second language comprehension, information processing, and performance in a simulated workplace condition, enabled us to predict benefits and consequences of using L1 and L2 in real-world situations as in Canada, where bilinguals have to carry out their daily activities and professional duties in a non-dominant language. The results of this study support adopting a cognitive (neuroscience) perspective to investigate the consequences of foreign language use in international business and organizational settings (e.g., Volk, Köhler, & Pudelko, 2014). Volk at el. (2014) reviewed models supporting the idea that foreign language processing depletes cognitive resources. They argued that implementing a common corporate language in multilingual workplaces, for instance, an English-only policy, may result in biased decision making, may reduce self-regulation, and consequently, may affect employees' performance. The practice of international organizations to mandate a common corporate language (e.g., English) to ease communication, and unite their employees with different language backgrounds, might negatively affect aspects of multilingual workers' performance. Further evidence for the influence of a non-native language on bilingual employees comes from a study by Neeley (2013), who investigated the effect of implementing English as a lingua franca, or a common language, on employee status loss in a French high-tech company. The findings of the study indicated that non-native speakers of English experienced status loss regardless of their degree of L2 proficiency. Fluency was correlated with language performance anxiety, job insecurity, and behavioral response– assertation, inhibition, or learning– to encounter with fellow native speaker coworkers.

The findings of the present study revealed that the type of task (enactment vs. verbal recall) influenced the recall accuracy in bilinguals and native speakers of English. Consistent with the results of previous studies that employed simple and dual-task paradigms in instruction-following tasks (e.g., Jaroslawska et al., 2018, Yang et al., 2014, Yang et al., 2016, Yang et al., 2018), participants in our study were consistently more accurate in acting out multi-step sequences of spoken instructions compared to verbally repeating the oral commands. This effect did not interact with language variables. It seems that the verbal recall task is more demanding for the cognitive system and relies on more cognitive resources. In previous studies of concurrent tasks, designed to disrupt the phonological loop, central executive, and visuospatial sketchpad components of working memory (Yang et al., 2016), the enactment of sequences of instructions was found to be less affected by the secondary tasks. The lower verbal recall accuracy of sequences of oral instructions suggests that the verbal recall is dependent on both the phonological loop and central executive components of working memory. Jaroslawska et al. (2018) assert that the action advantage provides support for Smyth and Pendleton's (1989, 1990) proposal that the availability of a motor buffer supports the temporary maintenance of movement trajectories and kinaesthetic representations. The advantage for action-based instructions is further linked to encoding in studies on patients with impaired memory. Individuals with Alzheimer's disease showed a deficit in free recall of sequences of instructions whereas they significantly benefitted from performing actions themselves at encoding (Charlesworth et al., 2014). Although the overall performance of following spoken instructions in patients with schizophrenia was impaired in Lui et al.'s (2018) study, mainly because of working memory impairments, the advantage of action-based instructions, both at the encoding and retrieval, was of equal effect size as for healthy controls.

In the current study, bilinguals' working memory capacity, as measured by the operation span task, was affected by the language of presentation. Bilinguals exhibited lower WM scores when they were doing the operation span task in L2. Considering that, so far, no theoretical accounts have considered separate WM components, such as the phonological loop and the visuo-spatial sketchpad, to be used for native language and the additional language (s) a bilingual has acquired, the differences in L1 and L2 WM scores can be attributed to the language of task (dominant vs. non-dominant). Whereas some researchers believe that WMC is language independent (e.g., Harrington & Sawyer, 1992; Osaka & Osaka, 1992;

Osaka et al., 1993), some others found that multi/bilinguals exhibit a larger WMC in their dominant language and that their L2/L3 WMC is affected by L2/L3 proficiency level (Coughlin & Tremblay, 2013; Service et al., 2002; van den Noort et al., 2006). In the present study, although L1 and L2 WMC were modestly correlated, $r = .22$, the language of task affected measures of WMC, resulting in a larger functional span in bilinguals' dominant language. We found a close relationship between individual differences in working memory capacity and following verbal instructions in bilinguals' and monolinguals' data. Participants with a higher WM span had higher recall accuracy in following sequences of spoken instructions in the enactment and verbal recall tasks. However, the interaction between WM span and the type of task was not significant, suggesting that the recall accuracy in the enactment and verbal recall tasks were dependent on working memory resources.

Furthermore, bilinguals exhibited lower accuracy and more errors in repeating non-words, especially longer pseudo-words, than their English native-speaking peers. However, there was no correlation between the scores of the non-word repetition task and the enactment and verbal recall tasks in bilinguals. In this study, bilingual participants were proficient in English at a near-native level. This is in line with previous findings that the relationship between language learning and phonological memory skills decreases with increasing language mastery (Gathercole, Willis, Emslie, & Baddeley, 1992; Service, 2012).

Contrary to our predictions, variables in bilinguals' language history did not affect their ability to follow instructions. A likely reason is that the bilinguals were a proficient homogenous group, who had already met English language proficiency requirements, at least an overall score of 6.5 out of 9 in the IELTS test or an overall score of 86 out of 120 in the TOFEL test. Another reason could be that the instructions in the current study were simple sentences with no syntactic complexity. Manipulating syntactic complexity or recruiting less skilled bilinguals might reveal the effects of language history.

2.4 Conclusions

The robust and specific advantage to bilinguals to have better recall accuracy in following spoken instructions in L1 compared to a fluent L2 demonstrated in this study indicates that linguistic and information processing and performing real-world tasks in a non-dominant language are demanding. The L2 input not only imposes additional computational and cognitive burdens on the processing system, but it also affects memory for serial order and the integration of information necessary to execute or verbally repeat sequences of spoken commands. Although bilingualism has been associated with many social and cognitive benefits (e.g., Bialystok, Craik, & Freedman, 2007; Calvo & Bialystok, 2014), input in a non-dominant language can influence processing (executive control) and storage (span/capacity) functions of working memory, resulting in lower efficiency in performing tasks that rely on sequential attention-based performance as in following spoken instructions in real-life situations. The fact that proficient bilinguals in the current study exhibited higher recall accuracy in L1 tasks shows that information processing and performance in a non-native language are cognitively costly and rely on superior language skills. This is confirmed by the results of the operation span task, in which the number of remembered words decreased and the reading/reaction times increased when stimuli were presented in L2. However, this study failed to find a strong effect of or correlation between memory measures and instruction-following ability although previous research points to the importance of working memory in recalling information presented in serial order. The variables in bilinguals' language history did not influence or interacted with instruction-following ability as our participants were homogeneous, with almost the same level of L2 proficiency.

The findings of this study have practical implications in the workplace, education, training, and organizational settings where a bilingual's second language is used as a common language (e.g., English) and the only medium of communication or instruction. As doing tasks in one's L2 seems cognitively taxing, international business and workplaces, where a bilingual has to use his non-dominate language, should train or spending money on improving bilingual employees' second language skills. Also, bilinguals might be given more time to do things in their non-dominant language.

3 Experiment Two

Memory for instructions in two languages: the effects of individual differences in working memory and language background variables

3.1 Introduction

Children and adults have to follow spoken or written instructions in various real-world situations, including daily activities, learning, training, and carrying out professional responsibilities. Previous research shows that following spoken instructions is affected by age and individual differences in working memory (WM), a limited capacity system that is responsible for the temporary maintenance and simultaneous processing of information (Baddeley, 2003, 2007, 2012, 2017; Baddeley & Hitch, 1974), in children (Engle et al., 1991; Gathercole et al., 2008) and adults (Allen & Waterman, 2015; Jaroslawska et al., 2016, 2018; Yang et al., 2016; Yang et al., 2014). For example, in the studies of Engle et al. (1991) and

Gathercole et al. (2008), older children and individuals with greater working memory capacity (WMC) performed better in encoding, retaining and acting out oral instructions. The role of WM in following spoken commands has further been shown in patients with impaired cognitive capacity (Charlesworth et al., 2014; Lui et al., 2018). Individuals with Alzheimer's disease (Charlesworth et al., 2014) and schizophrenia (Lui et al., 2018) have a deficit in recalling sequences of instructions.

Single and dual-task paradigms have been used in these studies. Further, either enactment or verbal recall, and secondary tasks during encoding or recall or both, have been employed. The type of task has been reported to influence the ability to follow oral instructions. Studies have reported the advantage of tasks involving acting out the instructions over verbally repeating them (Charlesworth et al., 2014; Lui et al., 2018; T. Yang et al., 2016). Participants' recall accuracy during encoding, at recall, or both, was higher when they enacted instructions either with real objects or in simulated conditions than when they were asked to listen to them and to orally recall them. To systematically block off theoretically proposed working memory components, such as the phonological loop or the central executive, simple tasks have been complemented with concurrent secondary tasks in dual-task paradigms. These are assumed to interfere with different working memory components. For instance, when participants are asked to perform or repeat sequences of instructions, a secondary task may put an extra burden on phonological loop or central working memory resources, resulting in both cases in lower recall accuracy in both verbal recall and enactment tasks (Yang et al., 2016).

Research shows that language and information processing in a non-native language can put a strain on the processing system, including on working memory, (Roberts, 2012), particularly in less-proficient bilinguals. Early work was reported by Dornic (1980). Recently, with increasing interest in the effects of internationalization, new studies targeting the effect of WM on a variety of tasks have emerged. A review of studies on the use of foreign language in international corporations (Volk, Köhler, & Pudelko, 2014) suggests that working in a non-native language depletes cognitive resources. This consequently hinders decision making and self-regulation in unbalanced bilingual employees. Unlike the first language (L1), second language (L2) processing is usually effortful and less automatic, especially, in less-skilled late L2 learners. Also, the non-dominant language has been reported to be more demanding for the computational and cognitive systems and to rely more on limited cognitive resources (Green, 1998; Linck, Osthus, Koeth, & Bunting, 2014; Meschyan & Hernandez, 2006; Perani & Abutalebi, 2005). Brain imaging research has shown that additional cortical areas are recruited by late low proficient bilinguals to process L2 input (Wang et al., 2007). Code-switching studies that manipulated L2 proficiency in unbalanced bilinguals (e.g., Costa & Santesteban, 2004) found significant switching costs when participants switched from a non-dominant to a dominant language. This was explained by assuming that more inhibition was needed to suppress L1 while participants were processing L2 input. Consequently, switching to L1 is assumed to require more resources because of the need to recover from the inhibition. If having less free WM resources results in less efficient processing, the extra load imposed by a non-dominant language can be expected to cause processing deficiencies, inaccurate mental representations, lower recall accuracy, and poorer task-performance among less-proficient L2 learners.

In addition to competition between languages, the degree of proficiency in a non-dominant language is a factor that can be assumed to play a role in language processing and task performance in bilinguals. The degree of L2 proficiency has been reported to interact with working memory measures and can influence processing and task performance in L2 (Coughlin & Tremblay, 2013; Hummel, 2009; Vejnović, Milin, & Zdravković, 2010). L2 proficiency can modulate WM span, resulting in larger spans for highly proficient bilinguals (Service et al., 2002; Vejnović et al., 2010). Greater proficiency can attenuate the burden imposed by a non-native language. In multilingual studies using neuroimaging techniques, L2 learners with low proficiency levels showed additional brain activity, mostly in prefrontal areas, in languages that they were not fluent in. They activated fewer neural substrates for sentence and discourse level processing in the left temporal lobe (Briellmann et al., 2004; De Bleser et al., 2003; De Bot & Jaensch, 2015; Perani et al., 1998; Perani & Abutalebi, 2005).

Another factor that has been reported to affect L2 learning, processing, and performance is the age of L2 acquisition. The effect of the age of the onset of acquisition (e.g., early vs. late) on language learning and processing has been widely studied (Archila-Suerte et al., 2015; Bloch et al., 2009; DeKeyser, 2013, 2017; Roncaglia-Denissen & Kotz, 2016; van den Noort et al., 2014; Wattendorf & Festman, 2008; Wattendorf et al., 2014). For example, the L2 AoA has been found to affect morphosyntactic processing, when proficiency is matched (Sakai et al., 2009), to influence the cerebral representation of language (Bloch et al., 2009) and the organisation of the cortical language network during sentence production (Wattendorf et al., 2014), as well as L2/L3 phonological processing (Archila-Suerte et al., 2015). In most of these studies, L2 processing, representation, and performance were influenced by individual differences in L2 acquisition. Early or late exposure to a non-native language resulted in differential performances in individuals who acquired or were exposed to the L2 later in life.

Furthermore, increased experience with L2 input and frequency of use can alter processing mechanisms, reduce L1 transfer effects, and ultimately result in native-like processing (Frenck-Mestre, 2002; Kroll et al., 2015; Pliatsikas & Marinis, 2013). The support for the positive role of exposure and experience with the target language come from studies of experience-based language processing and learning (Christiansen & Chater, 2016; Morgan-Short, Finger, Grey, & Ullman, 2012), statistical learning (Wells, Christiansen, Race, Acheson, & MacDonald, 2009), priming effects (Brandt, Nitschke, & Kidd, 2017), and training participants on infrequent structures (Deng et al., 2017; Hopp, 2016). For example, increased exposure to relative clause structures facilitates interpretation and processing speed of object relative clause structure than subject relatives (Wells et al., 2009). In addition, the attrition of an L1 in an L2 environment (see Schmid, 2016, for a review) and backward processing transfer, in which L2 parsing strategies are applied to process L1 input (Dussias & Sagarra, 2007), show that greater exposure to frequency of use of the target language can alter processing strategies in favor of the dominant language.

The Present Study

The present study follows from Experiment 1, aiming to replicate the superiority of L1 compared to L2 English in a learner group with a different L1. It again targets the ability to remember and follow oral instructions in a dominant and non-dominant language, a task that has ecological validity in many everyday contexts. The tested population further differed from the bilingual Persian L1 speakers of Experiment 1 in a number of aspects. We were interested in studying English L2 bilinguals whose first language, Mandarin, is not related to English. An added aim in Experiment 2 was to explore the role of variables in the bilinguals' language background, such as the age of onset of the L2 acquisition, functional L2 proficiency, amount of L2 exposure, and language use, as well as individual differences in WM measures. In Experiment 2, Chinese L1 bilinguals with intermediate English proficiency were tested in their first language and L2 English. The English proficiency of these Chinese-English bilinguals was poorer than that of the Persian-English bilinguals in Experiment 1 ($t_{(71)} = 3.87$, $p = <.001$), making it possible to better investigate effects of language proficiency. To our knowledge, this is the first study of this type to examine how L2 background variables and WM affect following multi-step sequences of verbal instructions in non-fluent bilinguals.

We hypothesized that the language of presentation (L1 vs. L2), language background variables, and individual differences in WM span and phonological memory would predict bilinguals' instruction-following abilities. The stimuli in the instruction tasks and memory assessments were presented in both languages of the L2 learners. The sequences of instructions were oral, mimicking possible situations in the workplace. Auditory input was thought likely to be more challenging for academic bilinguals and could be expected to tax more cognitive resources. Previous studies on listening comprehension in native and non-native languages, especially in adverse conditions, have revealed effects of the language of presentation, L2 proficiency, and WM capacity (Francis, Tigchelaar, Zhang, & Zekveld, 2018; Kilman, Zekveld, Hällgren, & Rönnberg, 2014; Sörqvist, Hurtig, Ljung, & Rönnberg, 2014; Van Engen, 2010). For instance, Kilman et al. (2014) investigated the effects of proficiency and WM capacity on English listening comprehension in L1 Swedish bilinguals. Participants listened to baseline and noise-masked Swedish and English sentences and repeated them. The results revealed that participants had more difficulties when the target speech was in the non-native language. L2 proficiency and L2 WM capacity were both correlated with speech perception in the non-native language. However, L2 proficiency was the stronger predictor of listening ability in the noise condition, suggesting that the

L2 WM measure may have been influenced by proficiency. A second study (Sörqvist et al., 2014) confirmed that participants with larger L2 WM capacity and higher proficiency are less susceptible to the effects of adverse conditions in L2 listening comprehension tasks. However, whether WM made a contribution to comprehension independently of L2 proficiency remained unresolved. Further, there could also be an interaction between language proficiency and working memory capacity, such that WM capacity could limit the ability to follow L2 instructions in lower proficiency L2 speakers more than in their higher proficiency peers.

We tested Chinese participants who were native speakers of standard Mandarin Chinese. Mandarin, like other varieties of Chinese, uses tones to distinguish words. Chinese is a Subject-Verb-Object (SVO) language with a pre-nominal adjective-modifier, for instance, 蓝色文件夹 ((the) blue folder). Like English, imperative sentences such as instructions, directions, and commands are constructed with imperative verbs, for example, **拿起**蓝色订书机 (gloss: *pick up blue stapler*: translation: *pick up the blue stapler*). Thus, a typical Chinese command/instruction would be:

“把红色订书机放到盒子里”

put red colour stapler machine to box in

‘put the red stapler into the box.’

3.2 Methods

3.2.1 Participants

We recruited forty Chinese-English bilinguals (24 females and 16 males; age *M* = 19.78, *SD* = 1.72, age-range = 18 – 26), who were native speakers of standard Mandarin Chinese. We posted the recruiting poster across the university campus, on the Facebook page of Chinese students, and had the poster displayed and the study announced at the beginning of a linguistics course attended by L1 Mandarin bilinguals. Table 3.1 summarizes the Chinese-English bilingual participants' language background information. Forty-nine native speakers of English (42 females and 7 males; age *M* = 20.47, *SD* = 4.50, age-range = 18 – 25; 79% born in Canada and 21% moved into the country before the age of 5) took part in the experiment as a control group. All participants were undergraduate students and were recruited through the SONA participant pool of the Department of Linguistics and Languages at McMaster University and received course credit for their participation. All participants had a normal or corrected-to-normal vision, normal hearing, and none reported diagnosed colour blindness. The study was cleared by the McMaster University Research Ethics Board.

Table 3.1 Summary of Chinese-English bilinguals' language background information.

Variable	M (SD)	range	Variable	M(SD)	Range
Age of arrival	17.10(1.69)	11.75 − 20	AoA	9.44(2.84)	4 − 15.5
Length of stay	2.82(2.14)	0.41 − 7.59	Proficiency	4.22(0.90)	1 − 6
MoA:			Daily use		
Classroom	8.40(2.48)	3.50 −13	L1	0.65(0.18)	0.20 − 0.93
Naturalistic	2.95(1.99)	0.41 − 7.59	L2	0.34(0.18)	0.05 − 0.80
Mental use			Preference		
L1	5.83(0.89)	3.83 − 7	L1	5.27(0.86)	3.33 − 6.66
L2	3.43(0.55)	1.50 − 5.33	L2	4.26(0.89)	2 − 6.16

Proficiency (1: very low – 6: native-like), language use in all daily activities (0.00 – 1), mental language use (1: never – 7: always), and language preference (1: never – 7: always) were assessed based on self-rating reports by bilinguals. Manner of acquisition (MoA), formal classroom instruction or naturalistic exposure to L2 in an L2 environment, the age of L2 acquisition (AoA), the age of arrival in an L2 country, and the length of stay in an L2 environment were reported in years.

3.2.2 Memory stimuli and tasks

We used the PowerPoint slides with the office-related objects and English stimuli in Experiment 1 as a starting point for this study. The oral instructions related to slides on a computer screen with a set of objects that could be dragged from one location to another on the screen. The target sentences consisted of four lists consisting of imperative sentences in the form of sets of five instructions. These were sentence-level materials by nature. The listeners were required to comprehend each set of five sentences and then execute or repeat the sequence of commands. There were 12 sequences of 5 instructions in each experimental list, 60 individual instructions in each combination of conditions, and, thus, an overall of 240 instructions for the four condition cells of the study. The pool of words used to create the sentences included 7 action verbs, 13 objects nouns, 7 types of documents, 5 colours, and 2 physical shapes (round/square). Colour and shape features created competitors for similar objects. Each sequence consisted of three simple instructions and one complex two-part instruction. Verbs such as "pick up, sign, print, and copy" and the connectives "and then" created the complex instructions involving two actions, for instance, *pick up the green ruler, and then put it into the square basket*. We avoided repeating the same verbs, objects, documents, colours or physical shape features in the same sequence of instructions. A typical example of a sequence of five instructions would be:

"*Move the yellow mouse onto the pad, print the budget report, and then put it into the blue folder, place the ruler into the round basket, and put the highlighter into the black tray".*

We controlled the length of the sequence of instructions. The average number of words in each sequence was 34 with a range of 33 – 35 words.

The English sentences were recorded by a female native speaker of English, and were digitized on a computer, using the Audacity software with a sampling rate of 44100 Hz. In prosody, we highlighted the distinctive features, for instance, the colours and shapes. Otherwise, the speaker was advised to follow the natural prosody of English. Each of the first four instructions was followed by 500 ms of silence, and there was a 1000-ms silence after the last instruction. A beep sound followed the last silence after instruction five, signaling the participant to commence the required action.

For Chinese tasks, the target sentences and experimental instructions were translated into Chinese by a native speaker of standard Mandarin Chinese. The translator was a graduate student at McMaster University. The translation was checked by two other native speakers for consistency. The target sentences were recorded by a female native speaker of Mandarin, Beijing dialect, from the Mainland, and were digitized on the computer. The correct use of tone was emphasized. The average number of characters in each sequence was 45 with a range of 42 – 49 characters. We followed the same recording and digitizing procedures as in the English task. Each participant received distinct lists for each language. We used Microsoft PowerPoint and SuperLab 5 to design and create enactment and verbal recall tasks. Bilingual participants filled out a customized language background questionnaire (see Appendix B) to report their demographic information and variables in their L2 language background.

3.2.3 Individual differences measures

To estimate participants' phonological memory skills, we used the English non-word repetition task (Archibald & Gathercole, 2006) (see Appendix C). Participants heard and said aloud 40 English-based non-words, recorded by a speaker of Canadian English. The items ranged from 2 to 5 syllables in length. The pseudo-words were digitally recorded onto a computer by a female native speaker of Canadian English, using Audacity software with a sampling rate of 44100 Hz. This task was scored by counting the number of correctly repeated syllables of the non-words.

To estimate working memory capacity, we used a task based on the English Operation Span Task (see OSpan task, Chapter 2) to measure native English speakers' WM span and Chinese bilinguals' L2 WM span. Participants read aloud and carried out sets of arithmetic operations each followed by a word that they had to memorize for later recall. For our version of the operation span task, we created a separate stimulus list for each language. Each list consisted of 15 sequences. Each sequence consisted of 2 to 6 arithmetic operations with 2 – 6 words for memory (see Appendix D for English stimuli). This resulted in 60 individual equations and 60 individual words for memory. For instance, a sequence of length two was as follows:

(7 × 8) – 4 = 52, (Y/N), BREAD

(6/3) + 9 = 11, (Y/N), MOTHER

RECALL

The English words to remember were mono- and bi-syllabic nouns taken from the MRC Psycholinguistic Database Output (Science and Technology Facilities Council, 2016). The words were checked for frequency, the number of syllables, letters, and phonemes. We also checked the words based on their familiarity, imaginability, and concreteness. For the Chinese language stimuli, we created and programmed a Chinese version of the OSpan task. The Mandarin words for memory were mono- and bi-syllabic words taken from Ho (2002) and a Frequency list of Chinese Characters in the Leeds University Online Corpus ("A collection of Chinese corpora and frequency lists," n.d.), based on 281 million words (tokens). The words were checked for frequency, the number of syllables, characters, and phonemes. Initially, we selected 120 nouns and then chose 60 target words as final stimuli after checking them based on familiarity, imaginability, and concreteness. As in English, we created 15 sequences, 60 individual equation-word pairs, for the Mandarin version of the OSpan task. Each sequence consisted of sets of 2 – 6 arithmetic operations and 2 – 6 words for memory. The order of items in a given sequence within a list was fixed; however, the sets within a list were counterbalanced so that participants could not predict the length of the upcoming sets. We counterbalanced the order of the assignment of OSpan lists for L1 and L2. We used SuperLab 5 software for programming and running the non-word repetition and OSpan tasks.

3.2.4 Design and procedure

The instruction experiment varied two within-subjects variables: the language of instruction presentation (L1 Mandarin vs. L2 English) and recall task (verbal recall vs. enactment). The order of items in a given sequence within a list was fixed. However, we counterbalanced the order of conditions and the assignment of stimulus lists for tasks and languages. Thus, each participant was assigned to two of the four stimulus lists in each language.

All bilingual participants were tested in two sessions, one in L1 and one in L2, with the language order counterbalanced between participants. We followed the same procedures as in Experiment 1 to test the Chinese bilinguals. The L1 sessions were facilitated by our LMBLab members who were native speakers of Chinese. Thus, all the communications and instructions were in Mandarin when the participants were attending an L1 session.

Memory for instructions was probed by either enactment or verbal recall. In the enactment task, the participant saw the pictures of objects, documents, and devices in each PowerPoint slide. They then double-clicked on a sound icon on the top left of each slide to listen to each sequence of instructions via headphones, with the goal of comprehending the instructions. They were instructed to memorize all the instructions in their serial positions in the sequence as the instructions were played through the headphones. In the enactment condition, the importance of performing the actions in correct sequence was emphasized. Upon hearing the beep sound after the last instruction in each sequence, participants had to use the computer mouse to immediately start performing the instructed actions on the computer screen, for instance, dragging a picture of a document onto a picture of a printer or moving an object into a container. Participants' performances were screen-recorded by the Quick Time Player software for data collection, scoring, and analysis.

In the verbal recall task, pressing a key on the keyboard would display the participants with a screen with pictures of office-related objects, documents, containers, and devices. At the same time, each sequence of auditory instructions related to the objects would begin playing on the headphones. The participants were advised to listen carefully as the audio file would be played only once. Upon hearing a beep sound after the last instruction, participants recalled the instructions in their serial order by saying them aloud. The importance of repeating the instructions in the correct order was emphasized. The pictures remained on the screen until participants pressed a key to proceed to the next sequence. Their voices were recorded by Audacity software for scoring and analysis.

The operation span task followed the enactment or verbal recall task and was presented in the same language as the L1 or L2 instruction task. Participants were asked to read aloud each arithmetic operation, perform the mental computation simultaneously and immediately press the specified Yes/No keys on the keyboard to decide if the provided answer was correct or incorrect. Upon pressing the Y/N key, they saw a word (a noun) in the upper case in red for 500 milliseconds on the next screen. They were asked to read it aloud as well. They continued doing the sets until they had gone through all items of a list and saw “RECALL” on the last screen. This required them to recall the words presented with each set of equations by saying them in the correct serial order. They had been instructed to say “BLANK” for each missed word, if they knew the sequential position of a word/words, but were not able to recall it/them. The order of the sets of different lengths was randomized so that participants might not be able to guess the length of each list. The experimenter sat at his desk and recorded the words the participants recalled and said aloud.

Reading times and arithmetic accuracy for each trial were collected by Superlab 5 software. Before running the experimental list, participants did a practice task consisting of two sequences of five instructions. The experimental instructions, the numbers, and the words were in either L1 or L2, depending on the language of the testing session. Bilinguals were asked to avoid doing the arithmetic computation in L1 if they were attending the L2 session.

The non-word repetition task was presented in the English session. After a practice trial, participants listened to individual non-words and repeated them aloud. The importance of careful listening and accurate repetition were emphasized. Participants' voices were recorded by Audacity software for scoring and analysis. In all tasks, the experimenter used the language of testing to greet, present experimental instructions and/or talk with the participants.

3.2.5 Statistical considerations

Two participants from the monolingual group and three participants from the Chinese-English bilingual group were excluded from data analysis because of unreliable or missing data. In the monolingual group, for one subject, the non-word repetition task had not been recorded for technical reasons and the other subject had skipped most items in the sequences in the enactment task. In the bilingual group, two participants did not show up for the second session and one subject had Cantonese as his first language.

The dependent variable was recall accuracy in enacting or verbally recalling the individual instructions. We modelled the recall accuracy as the probability of correct response for each instruction in a given sequence of five oral instructions. We used a sequence length of five instructions based on previous studies, similar to the Yang et al., (2016) study that investigated the involvement of working memory in following spoken instructions in adult native speakers of English. However, contrary to most similar studies that have relied on averaging the performance of each subject in each task and by employing ANOVA models for data analysis, we used the more powerful generalized linear mixed effects models with multiple variables and covariates. Given that the recall accuracy for individual instructions was bound by 0 and 1, violating the assumptions of linear regression models and ANOVA (see Jaeger, (2008) for discussion), we used the generalized linear mixed effects regression models (Baayen, 2008; Baayen et al., 2008) as implemented in the *lme4* packag*e* (Bates et al., 2014) and R Core Team, (2017) with a binomial distribution to perform a linear mixed effects analysis of the relationship between independent variables and recall accuracy for spoken instructions. This method of analysis allows to explore the effects of multiple factors and covariates while separately accounting for any variance contributed by participants and items (in this case, each individual instruction). As fixed effects in the models, we had the language of presentation (Chinese or English), group (monolingual or bilingual), the type of task (enactment or verbal recall) with scores on the complex working memory span, scores on phonological memory, and variables in bilinguals' language background as continuous covariates. In addition, we used the position of each instruction in a given sequence (1-5) and trial (running number of an item in a list, 1-60) as control variables. As random effects, we had intercepts for participants and items. Initially, we fitted each model with a maximal fixed effects structure and then removed the factors that did not significantly improve the model's performance. Where a model comparison was done, *p*-values were obtained by the likelihood ratio test of the full model including the effect in question against the model without the effect in question. Separate analyses were run to compare performance on English instructions between groups (monolinguals and bilinguals) and to compare within-group performance in L1 and L2 in the bilingual group.

The enactment of an individual instruction was scored correct if participants carried out the correct action on the said object or document in its correct sequential order position among the set of five instructions. Each individual instruction in the verbal recall task was scored correct if participants correctly repeated the instruction in its correct sequential position and with all details. Thus, incorrect or missed instructions received zero points. In the verbal recall task, any errors in the sequential order of instructions, recalling the wrong action verbs, objects or documents, containers or positions, or colour or shape features were the odds for getting an instruction correct and in the correct position out of five. However, as some actions verbs such as “place”, “put” or “move” required similar actions in the enactment task, the response was acceptable if participants recalled and said a synonym or similar action verb in the verbal recall task. Also, participants were not penalized if they forgot to recall the function words such as articles or prepositions in the verbal recall task.

The data collected from English and Chinese operation span tasks were scored based on the number of memorized words recalled correctly in serial order within each set of equation-word pairs. The number of memory words correctly recalled in their serial positions in each sequence was divided by the total number of words in that sequence (2 to 6). The scores were averaged for each participant and henceforth, the results of the operation span task will be reported as “working memory span (WM span)” or “working memory capacity (WMC)”.

The data of the English non-word repetition task were independently scored by two student members of the LMBLab. They were native speakers of English and majoring in Linguistics or Cognitive Science of Language programs.

Participants received partial credits based on the accurate repetition of the number of syllables in each repeated non-word. The scores were averaged for each participant, and henceforth, the results of the non-word repetition task will be reported as “phonological memory”.

3.3 Results

3.3.1 Enactment and verbal recall tasks

As descriptive data in Table 3.2 shows, the language of testing influenced bilinguals’ performance in both enactment and verbal recall tasks, with a disadvantage for tasks done in L2. Furthermore, all participants consistently were more accurate in acting out the sequences of instructions than verbally recalling them. The effect of the task appears to be present regardless of the language of presentation and group (see Table 3.2).

Table 3.2 Means and standard deviations of recall accuracy of acting out and verbally recalling sequences of five instructions.

		Enactment		Verbal Recall	
Group	Language	*Mean*	*SD*	*Mean*	*SD*
Bilinguals (*N* = 37)	Mandarin	0.59	0.49	0.44	0.50
	English	0.52	0.50	0.38	0.48
Monolinguals (*N* = 47)	English	0.68	0.47	0.45	0.50

For the bilinguals, one of the critical questions was if the language of the task, as a within-subjects factor, would influence the recall accuracy in acting out and verbally repeating sequences of oral instructions. We fitted the L1 and L2 data into mixed effects regression models to test the research hypothesis that presenting instructions in a non-dominant language would result in lower recall accuracy in bilinguals. The results of the models revealed a significant main effect of the language of presentation on recall accuracy in bilinguals. The language of testing significantly influenced bilinguals' recall accuracy in both acting out and verbally recalling sequences of spoken instructions, with a disadvantage for instructions presented in L2 (see Table 3.3 and Figure 3.1). There was also a significant main effect of recall task, with better performance for enactment than verbal recall. There were no significant interactions between the type of task (enactment vs. verbal recall) and the language of presentation (Mandarin vs. English), reflecting the recall accuracy advantage in both L1 enactment and verbal recall tasks. Participants also consistently performed better when acting out sequences of instructions than verbally repeating them in both L1 and L2 tasks, suggesting the advantage of carrying out the action than repeating the instructions verbally.

Table 3.3 Summary of the final logistic regression model of bilinguals' (*N* = 37) recall accuracy in following L1 and L2 instructions, reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number (out of 60) of the trial as counted from the beginning of the experiment. The reference level is indicated for categorical variables.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.6571
Participant	Intercept	1.0721
	Language = *Chinese*	0.6431
	Task = *Verbal recall*	0.5724

Part B: fixed effects

	Estimate	Std. Error	*z* value	*p* value
Intercept	0.0793	0.1881	0.422	0.6733
Position	-0.1765	0.0347	-5.087	<.001
Trial	-0.0018	0.0028	-0.638	0.5233
Language = *Chinese*	0.3034	0.1274	2.382	0.0172
Task = *verbal recall*	-0.7382	0.1179	-6.261	<.001
Language *Chinese** task *verbal recall*	0.0189	0.1002	0.188	0.8508

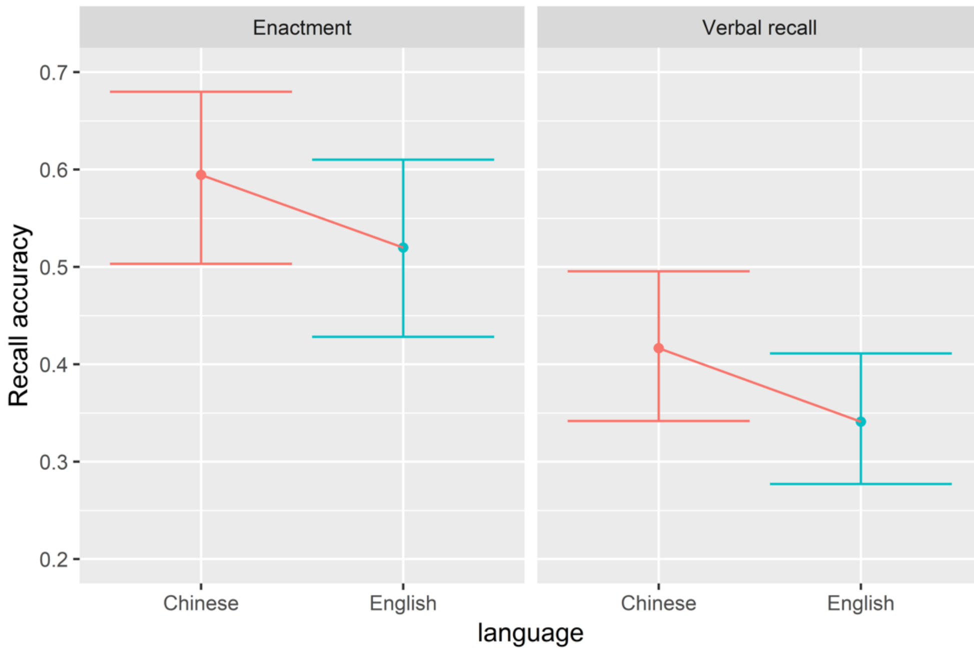

Figure 3.1 The effect of the language of presentation on recall accuracy in L1 and L2 tasks in Chinese L1 bilinguals. Error bars represent the standard error.

We further explored if the position of an individual item (instruction) in a given sequence of five instructions affected recall accuracy. The results of mixed effects regression models revealed that the serial position of an individual instruction significantly affected recall accuracy in L1 (b = -0.1838, SE = 0.0384, z = -4.789, p < .001, model not shown) and L2 (b = -0.1697, SE = 0.0429, z = -3.952, p < .001, model not shown), with recall accuracy being boosted by the primacy effect (see Figure 3.2). However, the running number of each trial in a list of 60 instructions did not significantly influence recall accuracy in either acting out or verbally recalling the instructions in L1 and L2. This suggests that there was little effect of learning during the task.

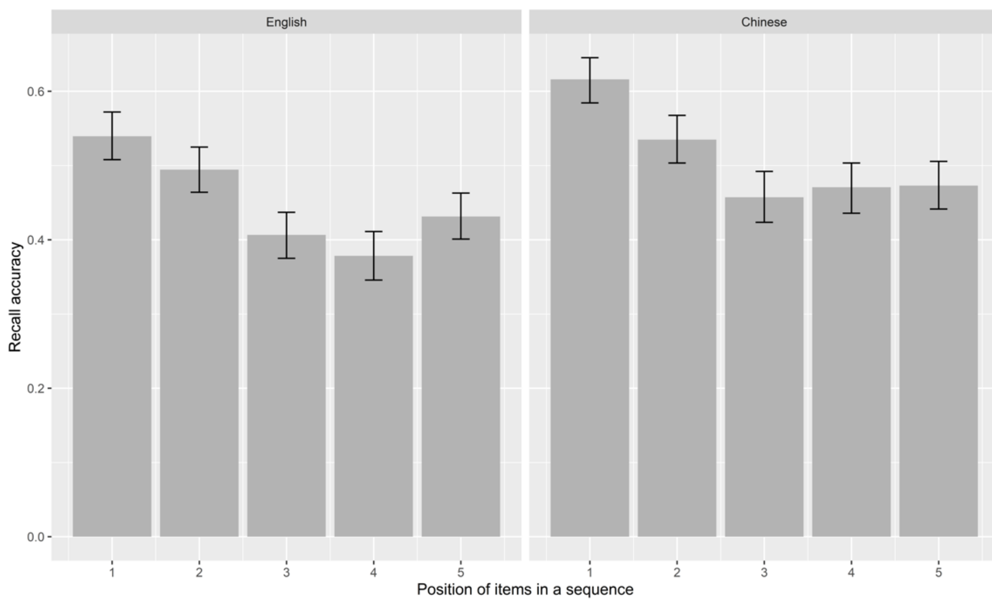

Figure 3.2 The effect of serial-order position of items in a given sequence of five instructions on recall accuracy in Chinese-English bilinguals. Error bars represent the standard error.

We also fitted the English monolinguals' data into generalized mixed effects regression models to examine the relationship between the type of the task and recall accuracy in following spoken instructions. The results of the models showed that there was a main effect of the type of task with significantly lower recall accuracy in the verbal recall task (b = -1.1623, SE = 0.1065, z = -10.918, $p < .001$, model not shown). Also, the serial position of instructions in each sequence of five instructions affected the recall accuracy (b = -0.2657, SE = 0.039, z = -6.834, $p < .001$, model not shown), showing a primacy advantage. Again, the running number of trials in a list of 60 instructions did not significantly influence recall accuracy in either acting out or verbally recalling the instructions.

A separate analysis was conducted to compare the performance of bilinguals and monolinguals in only English instructions. The results of the models (see Table 3.4), revealed that group as a between-subjects factor significantly affected the recall accuracy for sequences of English instructions. Bilinguals had lower recall accuracy in L2 English tasks, suggesting that they recalled fewer instructions than their monolingual peers in English tasks. However, there was a significant interaction between the type of the task and the group. The difference in the recall accuracy between the enactment and verbal recall tasks was much bigger for the monolingual than the bilingual group. In other words, although both groups were significantly poorer in the verbal recall task, the recall accuracy difference between the enactment and verbal recall tasks was smaller in bilinguals than in English monolinguals.

Table 3.4 Summary of the final mixed effects logistic regression model of the effect of group, bilinguals (*N* = 37) and monolinguals (*N* = 47), on the recall in English trials, reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number (out of 60) of the trial as counted from the beginning of the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name		Std. Deviation	
Item	Intercept		0.6982	
Participant	Intercept		0.8933	
	Task = *verbal recall*		0.5781	
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	0.9250	0.1453	6.364	<.001
Position	-0.2222	0.0359	-4.062	<.001
Trial	-0.0048	0.0029	-1.644	0.1003
Task = *verbal recall*	-1.1549	0.1047	-11.030	<.001
Group = *bilingual*	-0.8438	0.2077	-4.062	<.001
Task *verbal recall* * group *bilingual*	0.4081	0.1587	2.572	0.0101

Overall, the models showed that the language of the task (L1 vs. L2) significantly influenced bilinguals' performance in following multi-step oral instructions. Bilinguals' response accuracy significantly declined when the instructions were presented in their non-native language, especially in the enactment task. We also found that bilinguals had lower recall accuracy than their monolingual peers in English tasks. In line with the findings of Experiment 1

(Persian L1 study), the current results support the hypothesis that information processing in a non-dominant language is cognitively demanding. The results of the models are also consistent with the findings of Experiment 1 that carrying out sequences of instructions is easier than repeating them verbatim.

3.3.2 The effect of L2 background variables

Another question we explored in this study was whether variables describing the bilinguals' language history predicted their ability to remember and follow spoken instructions in their non-dominant language. To this aim, we added variables in the L2 learners' language history as predictors to the generalized mixed effects regression models modelling the bilinguals' L2 recall accuracy data. Participants and items were modelled as random effects. The fixed effects included participants' current age, education (undergraduate, Masters, or Ph.D.), gender, the age of arrival in an English-speaking country (Canada), length of residence in an English-speaking country (in years), manner of acquiring English as a second language (classroom instruction or natural exposure (in years)), mean onset of L2 age of acquisition in listening, speaking, reading, and writing skills, mean degree of proficiency in listening, speaking, reading, and writing skills (based on self-reported rating, ranging 1 – 6), L1 and L2 daily use (based on self-reported percentage), mental L1 and L2 use (based on self-reported rating, ranging 1 – 7), and L1 and L2 preference in academic and non-academic situations (based on self-reported rating, ranging 1 – 7) (see Table 3.1). An exploratory analysis of L2 background variables showed that there was a correlation between the L2 AoA and L2 proficiency, r = -.49 [95% CI: -0.71 – -0.20], suggesting that bilinguals who acquired English earlier in life rated themselves more proficient. However, the correlation between L2 proficiency and the manner of L2 acquisition was not significantly different, r = .20 [95% CI: -0.13 – 0.49] for classroom instruction and r = .17 [95% CI: -0.16 – 0.47] for the naturalistic exposure, suggesting that our sample of Chinese bilinguals had similar manner of L2 acquisition.

The results of the mixed effects logistic regression models showed that the mean onset of the age of acquisition in all L2 skills was a strong predictor of recall accuracy in following multi-step verbal commands in L2 tasks (see Table 3.5). As Figure 3.3 shows, L2 learners who acquired English as a foreign or second language later in life had significantly lower recall accuracy in following L2 oral instructions. As seen in Table 3.5 and Figure 3.3, the onset of L2 age of acquisition interacted with the type of task, with higher recall accuracy in the enactment than the verbal recall task for early but not late bilinguals.

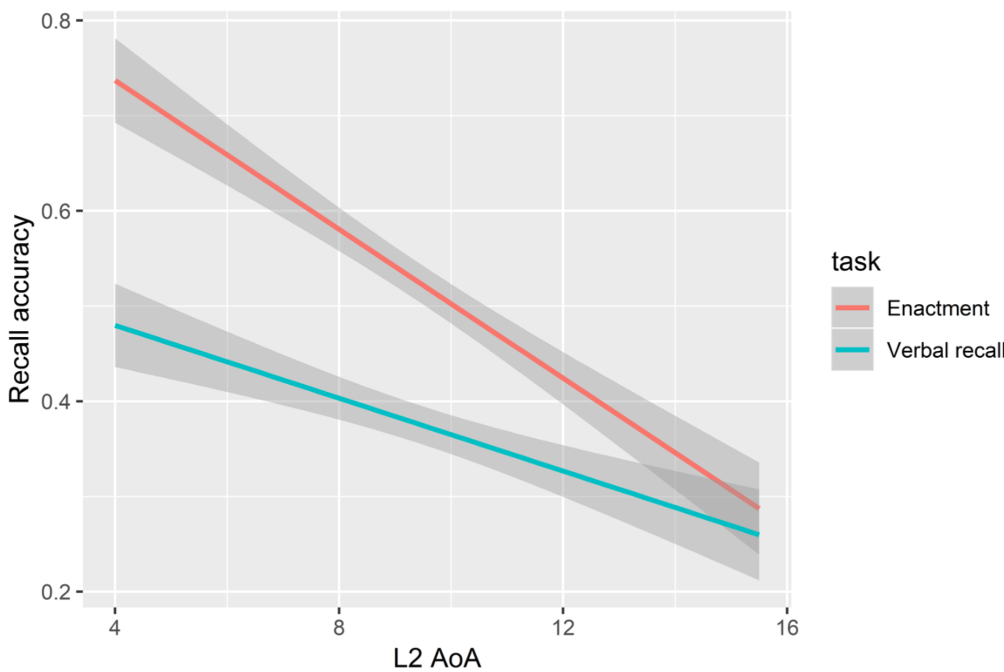

Figure 3.3 The relationship between the L2 AoA and the recall accuracy in following L2 instructions in Chinese-English bilinguals.

Table 3.5 Summary of the final logistic regression model of the effect of L2 age of acquisition on L2 instruction recall accuracy in bilinguals (*N* = 37), reported as the regression coefficient estimates, standard errors, z values, and p values. Position refers t

Part A: random effects				
	Name		Std. Deviation	
Item	Intercept		0.7683	
Participant	Intercept		0.8429	
	Task = *verbal recall*		0.5285	
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	2.1027	0.5277	3.984	<.001
Position	-0.1698	0.0430	-3.951	<.001
Trial	-0.0034	0.0035	-0.970	0.3318
Task = *verbal recall*	-1.5769	0.4048	-3.896	<.001
L2 AoA	-0.2136	0.0535	-3.994	<.001
Task *verbal recall* * L2 AoA	0.0844	0.0417	2.024	0.0429

We also explored the interaction between the L2 AoA and the exposure to L2. The result of the final model indicated a significant interaction between L2 AoA and daily exposure to L2 input and use, (b = -0.8487, SE = 0.2570, z = -3.302, p <.001, model not shown). Thus, bilinguals who acquired an L2 later in life were using L2 less frequently, and were more likely to have lower recall accuracy in following L2 instructions. The interaction between L2 AoA and the context of L2 acquisition was marginal only for the naturalistic exposure not the classroom instruction, (b = -0.0379, SE = 0.0222 z = -1.710, p = 0.0874, model not shown).

We further investigated if the language of task interacted with L2 AoA. We fitted instruction recall accuracy data of L1 and L2 tasks and L2 AoA into a series of mixed effects models. The results of the final model predicting the interaction between L2 AoA and the language of task showed that the interaction between these two factors was not significant, (b = 0.0556, SE = 0.0448, z = 1.241, p = 0.2147, model not shown).

The results of logistic regression models also showed that the degree of L2 proficiency was another critical factor that affected recall accuracy in following sequences of spoken instructions in a non-native language in Chinese bilinguals (see Table 3.6). As the results of the final model in Table 3.6 and Figure 3.4 show, more skilled bilinguals acted out and verbally recalled more sequences of instructions than their less-proficient peers. However, L2 proficiency did not significantly interact with the type of task (b = -0.0680, SE = 0.1749, z = -0.389, p = 0.6974, model not shown), suggesting that performing both the enactment and verbal recall tasks was affected by the degree of L2 proficiency. The result of the likelihood ratio test, X^2 (1) = 0.1575 p = 0.6915, showed that the interaction model was not necessarily a better fit. Therefore, we are reporting the results of the simpler model in Table 3.6.

Table 3.6 Summary of the final logistic regression model of the effect of L2 proficiency on L2 instruction recall accuracy in bilinguals ($N = 37$), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number (out of 60) of the trial as counted from the beginning of the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	0.7676		
Participant	Intercept	0.9527		
	Task = *verbal recall*	0.5807		
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-1.8565	0.8456	-2.195	0.0281
Position	-0.1697	0.0429	-3.952	<.001
Trial	-0.0034	0.0035	-0.972	0.3310
Task = *verbal recall*	-0.7833	0.1218	-6.428	<.001
L2 proficiency	0.5077	0.2164	2.347	0.0189

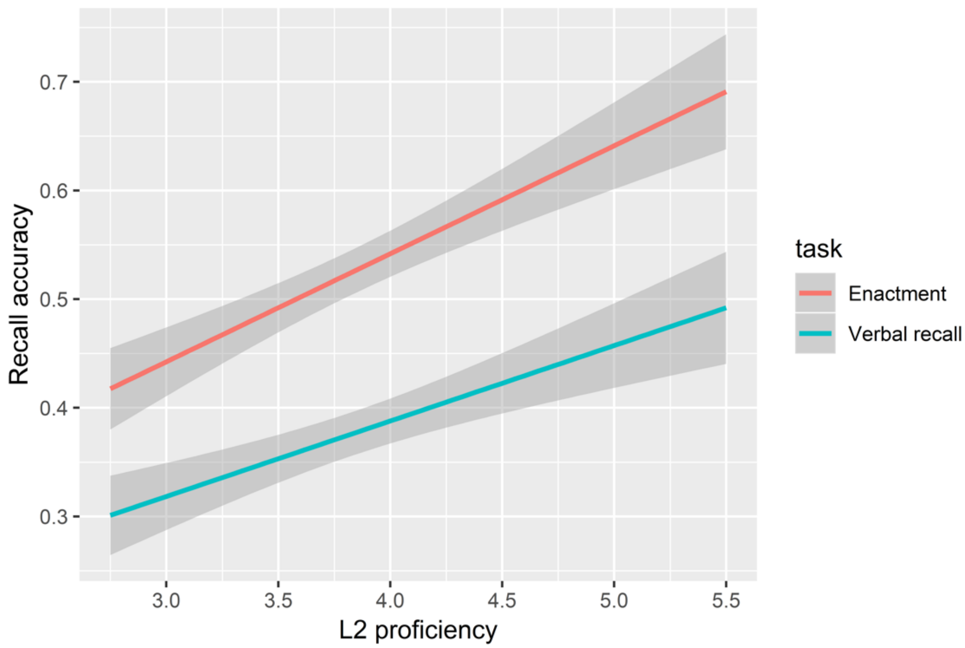

Figure 3.4 The relationship between L2 proficiency and recall accuracy in following L2 instructions in Chinese-English bilinguals.

We found no significant effects of other factors in bilinguals' language background, including age, education, gender, the age of arrival, length of residence, manner of L2 acquisition, daily L1 or L2 use, or L1 or L2 language preference, on recall accuracy in the enactment and verbal repetition tasks. Furthermore, these elements did not significantly interact with the type of task.

In sum, the models showed that L2 AoA and functional proficiency were two strong elements that influenced recall accuracy in following sequences of L2 instructions in Chinese L1 speakers. Bilinguals who had acquired English as a foreign or second language earlier in life or mastered the skills in their non-dominant language better had superior performance compared to late or less-proficient bilinguals. Interestingly, we did not observe a comparable significant relationship between factors in L2 language history and instruction-following ability in the more advanced Persian-English bilinguals in Experiment 1. Even, in the current study, we did not find any reliable or significant relationship between other elements in the bilinguals' language history and recall accuracy in following L2 instructions.

3.3.3 The effect of working memory capacity: The Operation Span Task

Another research question was to what extent individual differences in working memory capacity influence memory for instructions. To this aim, we examined the relationship between working memory and following sequences of oral instructions in bilinguals and monolinguals. We hypothesized that participants with larger working memory capacity would be better at recalling sequences of multi-step oral instructions. We further investigated if the language of the complex span task, operation span, influences WM span, assessed separately in L1 and L2.

As descriptive data in Table 3.7 suggest, the language of testing affected bilinguals' WM span. Bilinguals had better WM span scores in the Chinese complex span task than the English complex span task. The average arithmetic accuracy in both bilingual (L1 and L2 OSpan tasks) and monolingual groups was well above chance, suggesting that participants were actively engaged in the processing aspect of the complex span task, i.e., computing the output of the arithmetic operation.

Table 3.7 Summary of WM span in bilinguals (*N* = 37) and monolinguals (*N* = 47), reported as recall accuracy scores: proportions averaged across items, arithmetic accuracy, and mean RT of reading out equations and making verification decisions in milliseconds

Group	**Language**	**WM span (SD)**	**Arithmetic accuracy (range)**	**Mean RT (ms)**
Bilinguals	Chinese	0.73 (0.32)	0.92 (0.45-1)	6060
	English	0.60 (0.33)	0.91 (0.58-98)	8567
Monolinguals	English	0.66 (0.32)	0.89 (0.55-1)	7103

An independent-samples t-test was conducted to compare the scores of WM span in L1 and L2 conditions. The results showed that there was a significant difference in the scores for L1 WM span (*M* = 0.73, *SD* = 0.32) and L2 WM span

(*M* = 0.60, *SD* = 0.33) conditions, $t_{(72)} = 4.10$, $p < .001$. Chinese L1 bilinguals recalled fewer words from memory in the correct serial order in English trials than in Chinese trials (see Figure 3.5). These results suggest that bilinguals' functional working memory can be affected by the language of task. Despite the differences in the scores of L1 and L2 complex span tasks, L1 and L2 WM spans were highly correlated, $r = .66$ [95% CI: 0.42 – 0.81].

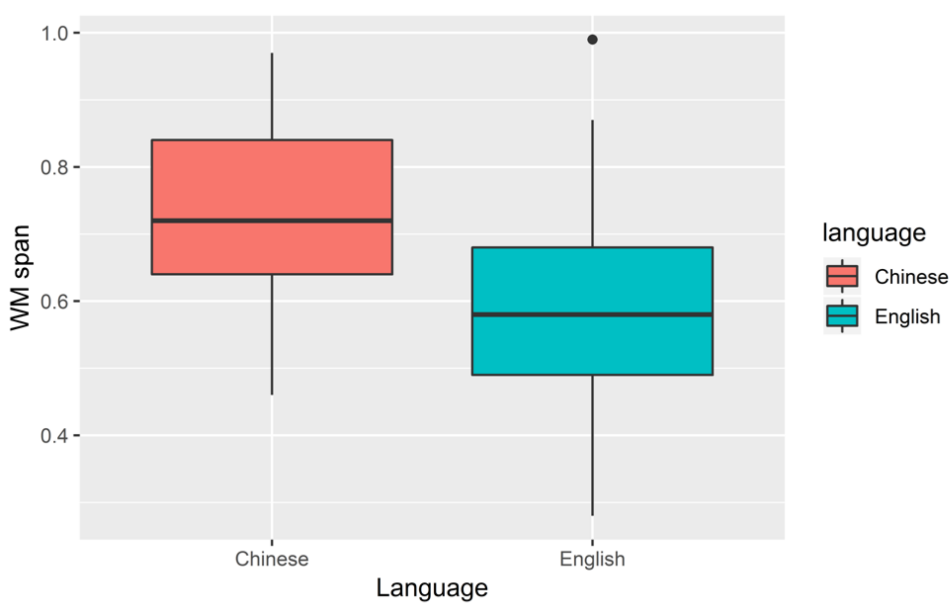

Figure 3.5 The effect of the language of presentation (L1 vs. L2) on the WM span in Chinese bilinguals.

However, in a model of English trials, comparing monolingual and Chinese L1 bilingual speakers, the result of an independent-samples t-test revealed that the difference between the means of WM span in bilinguals (*M* = 0.60, *SD* = 0.33) and English monolinguals was marginal (*M* = 0.66, *SD* = 0.32), $t_{(82)} = -1.96$, $p = 0.0544$. The mean score for the L2 WM condition was lower than the English monolinguals' WM condition.

We further examined the hypothesis that individual differences in working memory capacity would affect recall accuracy for oral instructions, with an advantage for individuals with a greater WM span having recalled more instructions. We entered the recall data of the L1 and L2 instruction-following tasks and the scores of L1 working memory span into a series of generalized mixed effects regression models. The results of the final model showed that WM span affected instruction recall with an advantage for participants with a higher WM capacity (see Table 3.8 and Figure 3.6). We did not find a significant interaction between WM span and the type of task, suggesting that the recall accuracy in both enactment and verbal recall tasks was correlated with WM scores. We further investigated if there was an interaction between language and WM span. The result of the final model showed that the interaction between language of task and WM span was not significant, ($b = -0.0887$, SE = 0.8756, z = -.0101, $p = 0.9193$, model not shown). As the language and WM interaction did not improve the model, we excluded it from the final model reported in Table 3.8.

Table 3.8 Summary of the final logistic regression model of the relationship between instruction recall accuracy and L1 WM span in bilinguals (N = 37), reported as the regression coefficient estimates, standard errors, z values, and p values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number (out of 60) of the trial as counted from the beginning of the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	0.6568		
Participant	Intercept	0.9971		
	Language = *Chinese*	0.6445		
	Task = *verbal recall*	0.5478		
Part B: fixed effects				
	Estimate	Std. Error	z value	p value
Intercept	-2.0520	0.8954	-2.292	0.0219
Position	-0.1765	0.0347	-5.087	<.001
Trial	-0.0018	0.0028	-0.638	0.5236
Language = *Chinese*	0.3130	0.1172	2.672	0.0076
Task = *verbal recall*	0.1703	0.5730	0.297	0.7663
WM span	2.8988	1.1962	2.423	0.0154
Task = *verbal recall* * WM span	-1.2270	0.7678	-1.598	0.1101

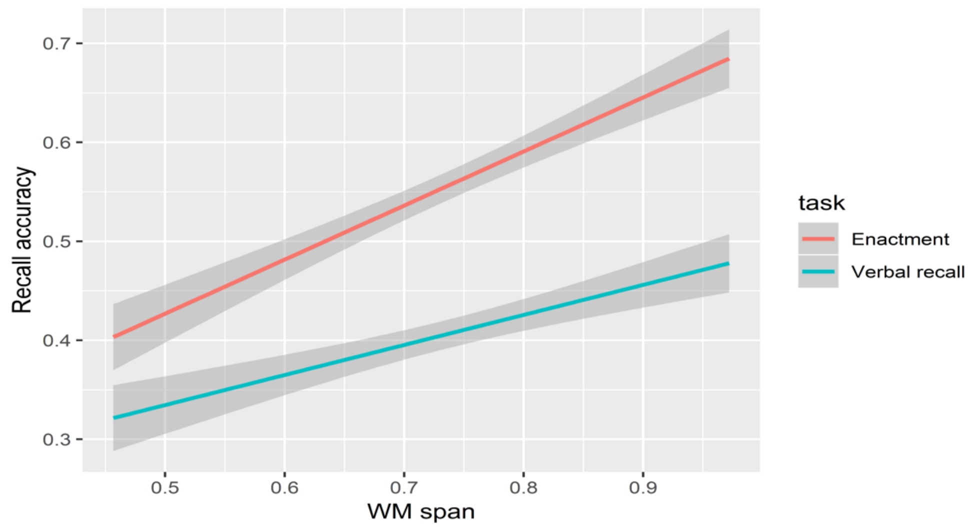

Figure 3.6 The relationship between L1 WM span and instruction recall accuracy in L1 and L2 tasks in Chinese-English bilinguals.

In a separate model, we explored the relationship between WM span and recall accuracy in monolinguals' data. As figure 3.7 shows, the WM span was correlated with recall accuracy in both the enactment and verbal recall tasks in monolinguals. Thus, participants with larger working memory capacity had a higher recall accuracy in following sequences of verbal instructions. The interaction between the WM span, the type of task, and language was non-significant (see Table 3.9).

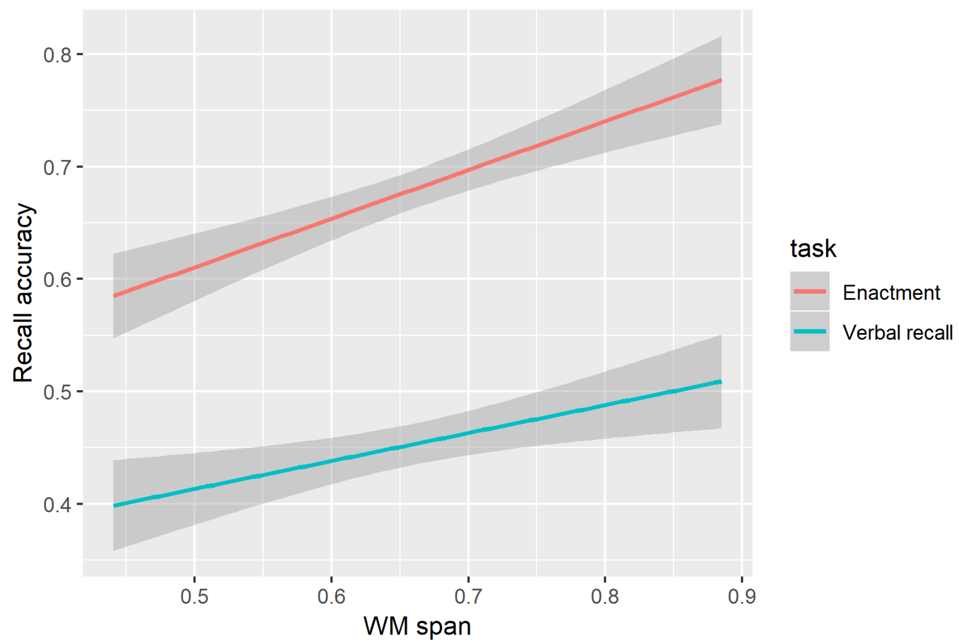

Figure 3.7 The effect of WM span on recall accuracy in the enactment and verbal recall tasks in the monolingual group.

Table 3.9 Summary of the final logistic regression model of the relationship between instruction recall accuracy and WM span in monolinguals (*N* = 47), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number of trials over the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	0.6995		
Participant	Intercept	0.7451		
	Task = *verbal recall*	0.5768		
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-0.5773	0.7113	-0.812	0.4170
Position	-0.2656	0.0389	-6.836	<.001
Trial	-0.0060	0.0032	-1.895	0.0580
Task = v*erbal recall*	-0.4786	0.6389	-0.749	0.4539
WM span	2.2947	1.0656	2.153	0.0313
Task *verbal recall** WM span	-1.0381	0.9593	-0.108	0.2792

Overall, the results of the models supported the hypothesis that WM span is correlated with the ability to remember and follow oral instructions. In both bilingual and monolingual groups, individuals with a greater WM span were able to both act out or verbally repeat a greater number of oral instructions. The effect of WM span was present regardless of the type of the task. Interestingly, although L1 and L2 WM spans were correlated, bilinguals' WM span was influenced by the language of testing. Bilinguals exhibited lower WM span scores when the stimuli were presented in the non-dominant language.

3.3.4 The influence of phonological WM: The Non-word Repetition Task

First, we compared the mean of phonological memory scores of bilingual and monolingual groups. The result of an independent-samples t-test revealed that Chinese L1 bilinguals (M = 0.85, SD = 0.06) had lower phonological memory scores than monolinguals (M = 0.97, SD = 0.05), $t_{(82)} = -11.23$, $p < .001$. As Figure 3.8 shows, bilinguals exhibited significantly lower accuracy in repeating English-based pseudo-words in the non-word repetition task than their monolingual peers.

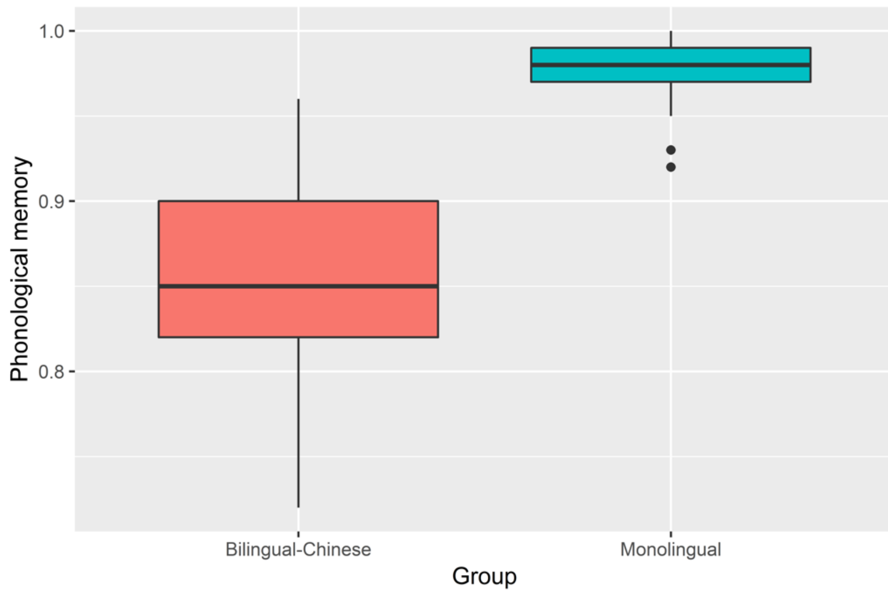

Figure 3.8 Comparison of phonological memory scores in Chinese-English bilinguals and English monolinguals.

We further investigated if phonological working memory affected recall accuracy in following spoken English instructions in bilinguals. We fitted the data of the English enactment and verbal recall tasks and the phonological memory scores from the non-word repetition task into a series of generalized mixed effect regression models. We tested the hypothesis that there was a relationship between phonological memory and the ability to remember and follow oral commands. The results of the final model predicting recall accuracy for English instructions revealed that the main effect of phonological memory was significant (see Table 3.10). As Figure 3.9 shows, bilinguals with larger phonological memory were more accurate in following multi-step oral instructions. Phonological memory interacted with the type of task with lower recall accuracy in the verbal recall task for participants with smaller phonological memory.

Table 3.10 Summary of the final logistic regression model of the relationship between instruction recall accuracy and phonological memory in bilinguals (*N* = 37), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. *Position* refers to the serial position of an instruction in a set of 5. *Trial* refers to the running number of trials over the experiment. The reference level is indicated for categorical variables.

Part A: random effects				
	Name		Std. Deviation	
Item	Intercept		0.7681	
Participant	Intercept		0.9580	
	Task = *verbal recall*		0.5361	
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-5.2419	2.4324	-2.155	0.0312
Position	-0.1696	0.0430	-3.948	<.001
Trial	-0.0034	0.0035	-0.972	0.3312
Task = *verbal recall*	2.5514	1.6742	1.524	0.1275
Phonological memory	6.2413	2.8432	2.195	0.0281
Task *verbal recall** phonological memory	-3.9003	1.9523	-1.998	0.0457

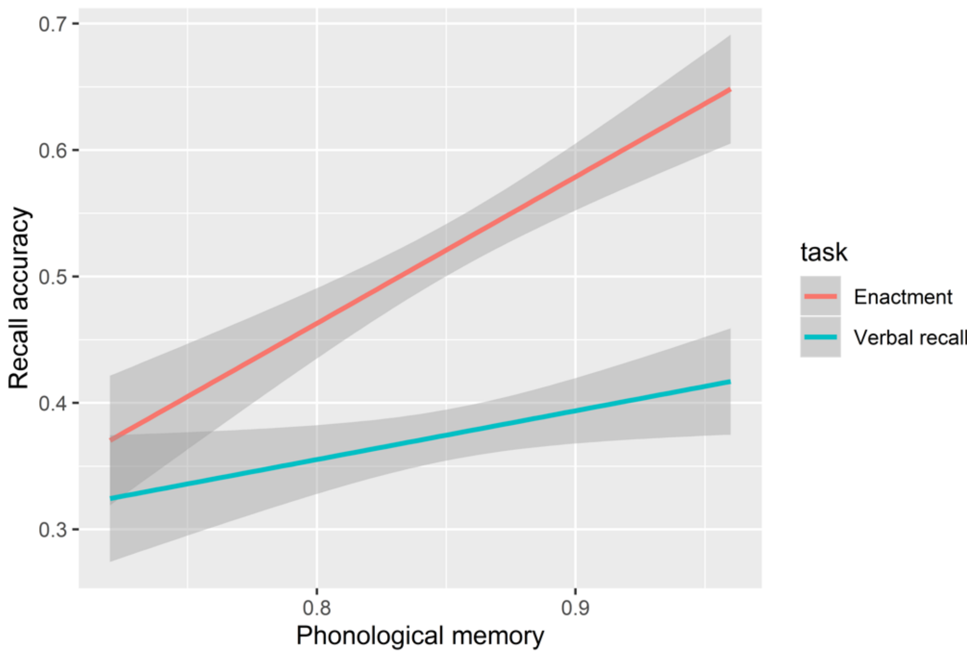

Figure 3.9 The effect of the phonological memory on recall accuracy of English instruction in bilinguals.

Taken together, phonological memory seems to be an influential factor in following sequences of oral instructions, especially in less-proficient bilinguals. We did not find such a significant relationship between recall accuracy and phonological memory in skilled Persian-English bilinguals in the first study.

3.4 Discussion

The present study is one of the first empirical investigations of following spoken instructions in bilinguals. We investigated the effect of the language of presentation and type of task to find out whether these factors affect recall accuracy in comprehending, remembering and executing sequences of multi-step real-world oral instructions. The experimental group consisted of Chinese-English bilinguals with intermediate L2 skills. In addition to the experimental variables, we tested variables related to individual differences in the bilinguals' L2 background, functional working memory and phonological memory as possible moderators of the potential effects.

The results of the mixed effects regression models comparing bilinguals' performance in L1 and L2 tasks revealed that the language of presentation (Chinese vs. English) affected bilinguals' instruction-following ability. Bilinguals' response accuracy significantly declined when the enactment and verbal repetition tasks were presented in their non-dominant language, with fewer instructions performed or verbally recalled in correct serial positions within a set in L2. Furthermore, the language of presentation affected bilinguals' functional working memory scores, as measured by the complex span task. Although bilinguals' L1 and L2 WM spans were correlated, bilinguals exhibited a larger working memory span when the stimuli were presented in their native language. The effect of the language of presentation on WM span can be the result of the additional computational and cognitive demands of processing L2 input or performing a task in L2 because of the strain a non-native language puts on the processing system (Roberts, 2012).

The degree of proficiency and other variables in bilinguals' language history can be expected to affect processing difficulties in L2, exhibit different L1/L2 brain activations, and cause inefficiency in performing real-world tasks in L2. The results of mixed effects models exploring the relationship between factors in bilinguals' language history and recall accuracy in following spoken instructions showed that L2 age of acquisition and degree of general proficiency were two important factors that greatly influenced L2 instruction-following ability. The L2 AoA highly affected bilinguals' recall accuracy in the enactment and verbal repetition tasks. Bilinguals who acquired English earlier in life, e.g., between the age of 4 – 10, had higher recall accuracy in following verbal instructions in a non-dominant language. The significant negative correlation between the onset of L2 AoA and the number of instructions correctly performed or verbally repeated reveals the linguistic and cognitive consequences of an additional language acquired later in life. However, there was a significant interaction between L2 AoA and daily L2 use, and a marginal interaction between L2 AoA and the manner of acquisition (naturalistic exposure). This suggest that the main effect of L2 AoA might be related to the amount of exposure and the context of language acquisition. The results are consistent with studies investigating the relationship between the onset of L2 AoA and L2 phonological processing (e.g., Archila-Suerte et al., 2015), morphosyntactic processing (Sakai et al., 2009), the cerebral representation of language (e.g., Bloch et al., 2009), as well as the organisation of the cortical language network during sentence production (e.g., Wattendorf et al., 2014). For example, in an fMRI study, Bloch et al. investigated the effect of L2 AoA on the cerebral activation during language production in proficient multilinguals. The results of their study revealed that the age of L2 acquisition correlated with and modulated the variability of brain activation in all three studied languages, English, (Swiss) German, and French, with low variability in early multilinguals and higher variability in late multilinguals. There was an increase in the individual variation of local cerebral activation in Broca's and Wernicke's areas in the later acquired languages, regardless of the typological differences between the acquired languages.

The degree of L2 proficiency was another critical variable in bilinguals' language history that influenced following L2 verbal instructions. Proficient bilinguals had higher recall accuracy in L2 enactment and verbal repetition tasks, suggesting that greater L2 proficiency results in better processing and more efficiency in performing real-life tasks in intermediate-level bilinguals. It has been argued that bilinguals with low L2 proficiency show additional brain activity, mostly in prefrontal areas that are associated with working memory, and activate less neural substrate (Green, 1998; Perani & Abutalebi, 2005), which support the impact of the degree of general L2 proficiency on L2 processing and performance. Our results are also consistent with studies that found that proficient L2 learners are more efficient in processing, have more accurate responses and mental representations, and can demonstrate native-like processing (Bel et al., 2016; Keating, 2017; Rosselli, Ardila, Lalwani, & Vélez-Uribe, 2016; Tanner, Inoue, & Osterhout, 2014; van Hell, 2010).

A question of interest was whether language effects on memory for instructions would be moderated by individual differences in WM capacity. The results of the final mixed effects regression models exploring the relationship between individual differences in working memory and instruction recall accuracy revealed that working memory capacity influenced the ability to comprehend and follow verbal commands. In both Chinese L1 bilinguals and English monolinguals, participants with a higher WM span demonstrated higher recall accuracy in acting out or verbally repeating sequences of oral instructions. In both groups, we did not find a significant interaction between WM span and the type of task, suggesting the main effect of working memory on recall accuracy in both the enactment and verbal recall tasks. In bilinguals, the interaction between the language of task and WM span was not significant. This suggests that there was the main effect of working memory on recall accuracy in both L1 and L2 tasks, and that language did not have a moderation effect. The results are consistent with previous instruction-following studies that reported the involvement of WM resources in following written or oral instructions or directions in children and adults (e.g., Engle et al., 1991; Gathercole et al., 2008; Jaroslawska et al., 2016, 2018; Yang et al., 2014, 2016). The relationship between the complex WM span and the ability to perform or orally repeat sequences of verbal instructions supports the hypothesis that difficulties in following spoken instructions can be caused by an extra burden put on the cognitive system. We think that the processes underlying the complex span task, i.e., simultaneous processing and temporarily holding on to the critical information for later recall, are also active during the instruction-following tasks. Thus, participants must encode each individual instruction, integrate the information with WM content, try to remember its position in the sequence, and then recall information from the whole sequence to execute or verbally repeat all the instructions. Therefore, the processes used for the temporary storage and manipulation of information in WM, are involved in following spoken instructions as well. However, besides encoding, processing, storage, and retaining the sequences of instructions, the successful performance of an instruction requires the inhibition of enacting or verbally repeating the previous instructions. The inhibition process, along with other above processes, may overload WM resources, making the instruction-following tasks demanding for the cognitive system.

Phonological memory seems to be an influential factor in the performance of less-proficient bilinguals. Phonological memory significantly affected L2 instruction recall accuracy in bilinguals. Bilinguals who had higher scores in correctly repeating English-based pseudo-words were able to remember and follow more L2 instructions. Phonological memory also interacted with the type of task. Bilinguals who had lower phonological memory were not able to remember and follow as many instructions compared to bilinguals with greater phonological memory. It seems that the effect of phonological memory extends to adulthood, at least in less-skilled bilinguals.

A consistent finding in all analyses was that the enactment of instructions was easier than verbally repeating them. One explanation is that, in addition to the verbal representation, it can act as a second coding of information (Paivio & Desrochers, 1980). This could result in better performance than for verbal recall, which does not have necessitate encoding in a second modality. There might be also interference/competition for having to do verbal rehearsal and also do verbal production.

The relationship between the WM span and the language of presentation seems to be more complex than what would follow from L2 simply adding to the WM load. In our experiment, although there was a correlation between L1 and L2 WM spans, WM span was affected by the language of presentation. Bilinguals exhibited a greater WM span in L1 than L2. For example, the results of van den Noort et al. (2006) on a group of L1 Dutch, L2 German, and L3 Norwegian multilinguals revealed differences in performance in all three languages. Participants had larger functional WMC in the L1, followed by the L2, then L3. Because language learning, development, and processing are dynamic processes (de Bot, 2012; De Bot, Lowie, & Verspoor, 2007, Lowie & de Bot, 2015) and language skills can be improved by experience with the target language over time, WMC in L2/L3 may reach the level of L1 as the result of mastery in L2/L3. This is consistent with other findings that linguistic experience and proficiency can affect L2 WM span (Reichle et al., 2013; Service et al., 2002; van den Noort et al., 2006). Service, Simola, Metsänheimo, and Maury (2002) argue that with an increase in L2 proficiency, less working memory resources are consumed when performing target language tasks.

3.5 Conclusions

The findings of this study add to accumulating results that L2 processing and performance rely on a combination of cognitive factors. Better memory for instructions was seen in both verbal and enactment responses. Presenting input in a non-dominant language seems to put an additional burden on the cognitive system, especially in less-proficient bilinguals. Consequently, it affected their instruction-following ability. The better recall accuracy in following sequences of instructions in participants with a greater WM span and phonological memory demonstrated in the current study indicates the importance of available working memory resources to perform complex multi-step tasks. Individuals with poor working memory are more likely to fail in tasks that require the involvement of different WM functions and resources. The onset of the L2 age of acquisition and proficiency were the two most influential factors in bilinguals' language background that were found to affect the instruction-following ability in bilinguals. The robust advantage of acting out oral instructions compared to verbally repeating them shows that the enactment of the required actions can act as a second coding of information. Overall, following spoken instructions draws on different working memory resources and is affected by some variables in bilinguals' language background and the language of task (dominant vs. non-dominant). Individual differences in the WM span and phonological memory, the onset of L2 AoA, and the degree of L2 proficiency, and the language of task are good predictors of the instruction-following ability in bilinguals.

The findings of this study can be applied to real-world situations such as the workplace, education, training, and organizational settings where bilinguals' non-dominant language (e.g., English/French in Canada) is used as the only medium of communication, instruction or performing responsibilities.

4 Experiment Three

The impact of the linguistic complexity on following verbal instructions containing sequential temporal order

4.1 Introduction

The findings of studies in chapters 2 and 3, on following sequences of five linguistically simple English instructions by Persian and Chinese L1 bilinguals, revealed that the language of presentation (L1 vs. L2) significantly affected recall accuracy reflected in acting out and verbally recalling spoken instructions. Bilinguals' recall accuracy was lower in their non-native language than in the dominant language. However, the effects of other factors such as individual differences in WM span, phonological memory, and variables in language background were mixed. Individual differences in WM span significantly correlated with instruction-following ability in both proficient and less-proficient bilinguals and the monolingual group. Thus, participants with a greater WM span had higher recall accuracy in the enactment and verbal recall tasks. We did not find a significant interaction between the language factor (L1 vs. L2) and WM span in bilinguals, suggesting that working memory capacity influenced instruction recall accuracy in both L1 and L2 tasks. Also, the interaction between the language factor and variables in L2 background was not significant. However, the main effect of phonological memory was only significant in the Chinese L1 speakers who were less skilled in L2 English. Participants who had greater phonological memory scores in the non-word repetition task also had a higher recall accuracy in acting out and verbally recalling sequences of instructions in their non-native language. Thus, the effect of phonological memory was more robust when less-skilled bilinguals were engaged in L2 instruction-following activities. Likewise, the relationships between some variables in L2 language history, such as L2 age of acquisition (AoA) and proficiency, and memory for instructions were significant only in the less-proficient Chinese-English bilinguals but not in the more skilled Persian L1 bilinguals. The syntactic constructions of the instructions in Experiments 1 and 2 were simple. In the third study, we manipulated the linguistic complexity of instructions that involved the sequential temporal order of actions. Thus, the study in this chapter explored the role of conceptual complexity in remembering and acting out multi-step instructions. In particular, we manipulated the match versus mismatch between the order of action descriptions in verbal instructions on the one hand and in the action sequences they described on the other. Our main interest lay in whether this type of linguistic complexity interacted with the language of the instructions and various individual-differences variables in predicting recall accuracy in bilinguals.

The sequential order of the events in adverbial clauses containing *before* and *after* connectives may or may not be consistent with the factual order in which the events unfold in real-world situations. Temporal order connectives such as *before* and *after* aid the integration of information in main and subordinate clauses. They also signal how the events link together to construct a coherent mental representation. For instance, in two-clause sentences, *Before/After the linguist submitted the paper, the journal changed its policy* vs. *The journal changed its policy before/after the linguist submitted the paper*, event A may follow or precede event B and the conceptual order of the two events may or may not match the order of mention (presentation). Although the underlying temporal information remains unchanged, the surface structure can take two different forms, either matching or reversing the temporal order as signaled by the *before/after* connectives.

While in the matching condition the order of mention of the events is congruent with the conceptual (actual) order, the conceptual order of the events in the reverse condition does not match the order of mention of events. Thus, the matching condition presumably requires less processing, being associated with lower comprehension load as it follows the dominant expectation that events are described in the chronological order that they happen in the real world. In contrast, in the reverse case, successful comprehension of the input and the construction of a coherent mental representation requires additional discourse-level computations and, can, therefore, be assumed to rely on more cognitive resources. Because the human mental representation of the events corresponds to the chronological order, and events are represented in memory in chronological order, regardless of the order of presentation in discourse (Zwaan & Radvansky, 1998), any violations of the default expectation should be costly for the computational and cognitive systems. The order of clauses in the reverse expressions must be mentally re-arranged to construct a coherent mental representation. Previous research has revealed that reverse surface order affects recall accuracy (Clark & Clark, 1968; Clark, 1971), reading speed (Mandler, 1986), elicits a negative-going event-related potential (ERP) component (N400) over anterior sites (Politzer-ahles, Xiang, & Almeida, 2017) and causes processing and comprehension difficulties in children (Blything & Cain, 2016; Blything, Davies, & Cain, 2015) as well as aphasic and

Parkinsonian patients (Ansell & Flowers, 1982; Natsopoulos et al., 1991; Oron, Szymaszek, & Szelag, 2015).

Although temporal order clauses appear in children's speech by the age of three (Diessel, 2004), they can cause processing difficulties even in older children. Studies on children have associated difficulty in processing before/after two-clause sentences with developmental challenges in understanding the meaning of temporal connectives and mentally representing the clauses in the reverse case (Amidon & Carey, 1972; Blything & Cain, 2016; Blything et al., 2015; Mandler, 1986; Natsopoulos & Abadzi, 1986). As the linguistic knowledge of connectives in adults is robust, any difficulties in comprehending such complex structures in adult age can be assumed to reflect additional costs that mapping reverse ordered descriptions to event order representations imposes on the computational and cognitive systems.

Early studies on children (Amidon & Carey, 1972, Natsopoulos & Abadzi, 1986) revealed that children's reading speed and enactment accuracy were affected if events were presented with descriptions mismatching with their chronological order. In two narrative reading experiments, Mandler (1986) found that comprehension was faster when the order of mention matched the real order of the events. However, the results interacted with the type of relation between the events, with arbitrary temporal relations being less affected than causal relations by the match/mismatch condition.

In two behavioral and functional magnetic resonance imaging (fMRI) studies, Ye et al. (2012) investigated the comprehension of temporal connectives in "before'- and 'after'-initial sentences in German patients with Parkinson's disease. In after-initial complex sentences, the sequence of events matched the order of mention, "*After A, B*". In before-initial two-clause sentences, the actual order of occurrence was not consistent with the order of presentation, *"Before B, A"*. Compared to the control group, patients showed pronounced difficulties in understanding the temporal sequence of the events in the reversed 'before' sentences. The results of the fMRI study revealed that a set of cortical and subcortical areas were associated with the processing of temporal order, suggesting that additional computations were required for 'before' sentences in the reverse condition.

Taken together, the results of several online and offline studies indicate that reverse constructions affect processing and comprehension of temporal order clauses in children and adults. However, unlike children, adults have fewer difficulties in distinguishing the meaning of before/after connectives. They are able to correctly interpret temporal relations. However, also they are faster at processing and better at recalling expressions in which two events are ordered chronologically (Evers-Vermeul, Hoek, & Scholman, 2017). Contrary to the default matching construction, re-arranging the events in the reverse construction to create a mental event representation relies on additional computations and is more demanding for the cognitive system. In the present study, we explored the hypothesis that L2 adds an extra load to cognitive processing so that effects from cognitively demanding tasks, such as processing reverse order descriptions of events, are exacerbated. Furthermore, we examined if the linguistic complexity in temporal order constructions correlates with linguistic background variables in a group of skilled Persian-English bilinguals although such effects were not present for the simple instruction sentences in the first experiment.

The Present Study

The present study aimed to investigate the effect of the linguistic complexity of temporal order in two-clause sentences on the recall accuracy and performance of sequences of spoken instructions. Two groups were tested: a group of native speakers of English (henceforth, monolinguals) and a group of first-language (L1) Persian bilinguals, who listened to the instructions in their dominant and non-dominant language (English). Specifically, we were interested in constructions expressing sequential temporal relations (using e.g., *before* and *after* connectives) rather than a synchronous temporal relation (using e.g., *while*). To explore how the congruency of the event orders influences participants' recall and performance of verbal instructions, we manipulated the order of the events, so that the order of mention of the events in two-clause sentences either matched or mismatched the conceptual order. We further investigated if there was any relationship between individual differences in working memory capacity (WMC) or phonological short-term memory (PSTM) abilities and following semantically complex L2 instructions. Research shows that unlike L1 processing, which is highly automatic, processing input in a non-dominant language, especially in less proficient late L2 learners, is demanding for the cognitive system and relies on more cognitive resources (Green, 1998; Linck, Osthus, Koeth, & Bunting, 2014; Meschyan & Hernandez, 2006; Perani & Abutalebi, 2005). Therefore, for bilinguals, in addition to the linguistic complexity and individual differences in WMC and PSTM measures, we manipulated the language of presentation, i.e., the first language (L1) vs. the second language (L2). We also explored the influence of individual variables in the language background, such as the onset of the L2 age of acquisition (AoA) and the degree of L2 proficiency.

While having greater WM resources may result in better and more efficient processing, the extra load imposed by a non-dominant language, or language complexity, may cause processing deficiency, inaccurate verbal memory representation, and poor performance in less skilled bilinguals. In the current study, the experimental stimuli to remember were spoken instructions. Research has shown that auditory input is more challenging for bilinguals and can be expected to tax more cognitive resources than written input (Francis et al., 2018; Kilman et al., 2014; Sörqvist et al., 2014; Van Engen, 2010). The stimuli were complex two-clause imperative sentences consisting of a main clause and a subordinate clause, connected by *before* or *after* conjunctions. The sentences used either temporally *matching* or *reverse* constructions, and were presented in either in L1 or L2. We presented the sentences verbally and asked the participants to execute the instructions on a computer screen. We hypothesized that the linguistic complexity of the temporal order construction would affect comprehension, recall, and enactment of complex instructions. We also expected that there would be a relationship between working memory and phonological memory abilities and following complex instructions (Gathercole et al., 2008; Jaroslawska et al., 2016; Lui et al., 2018; Yang et al., 2017; Yang, Jia, Zheng, Allen, & Ye, 2018b). The language of presentation was hypothesized to influence bilinguals' recall accuracy and performance as has been reported, for example, in L1 and L2 arithmetic processing (McClain & Huang, 1982; Wang, Lin, Kuhl, & Hirsch, 2007), task-performance in L2 (Neeley, 2013; Volk et al., 2014), and L2 processing (Roberts, 2012). Further, we expected some language background variables to correlate with the recall accuracy in acting out of complex instructions in the L2 of the bilinguals. Presently, it is not clear how individual differences in working memory capacity and phonological short-term memory might interact with the alternative constructions in temporal clauses. Furthermore, how the language of testing and variables in an individual's language history affect processing of the complex temporal clauses in non-native speakers is of interest for a better understanding of the character of the extra load being imposed by time-reversed constructions. In particular, it is not clear if this type of load interacts with the expected L2 language load.

The bilingual participants were native speakers of Persian, a Subject-Object-Verb (SOV) word order language. In Persian, commands can be constructed either by adding particle *be* to the bare root of the verb (*e.g., be-gozar*, put) or without the particle in verb phrases containing light verbs like *kærdæn* (to do), (*e.g., print kon*, printing do: print). Persian complex temporal-order clauses are constructed by having temporal order connectives *qæbl ʔæz inke* (before from that:

before) or *bæʔd ʔæz inke* (after from that: after) at the beginning of the subordinate clause. Like English, the main or subordinate clause can occur as the first or second sentence in two-clause sentences. A typical complex command in Persian would be:

1. *qæbl ʔæz inke resid ra kopi kon-id, form e tæqaza*
before from that receipt OBJ-marker copying do-2 SG, form application

ra mohr be-zæn-id
OBJ-marker stamp IMP PTCL- hit-2 SG

‘before you copy the receipt, stamp the application form.'

We predicted that participants would have fewer correct responses in oral instructions containing conceptual temporal-order constructions and that this would interact with the language of presentation.

4.2 Methods

4.2.1 Participants

Thirty-five Persian-English adult L2 learners (15 females and 20 males; age *M* = 27.71, *SD* = 5.91, range = 19 – 45) participated in this study. They were compensated 15 $/hour and were recruited from undergraduate and graduate programs at McMaster University by posting the recruitment posters on the Facebook pages of Persian speaking students and the Iranian community and across the campus. The Persian-English bilingual participants filled out a customized self-report language background questionnaire (see Appendix B) to report their demographic and linguistic information. They also rated their L2 functional proficiency, L1 and L2 daily and mental use, and L1 and L2 preferences in different situations (see Table 4.1 for a summary of bilinguals' language background information). Forty-two native speakers of English (38 females and 4 males; age $M = 19.52$, $SD = 0.99$, range = 18 - 22; 88% born in Canada and 12% moved into the country before the age of 5) took part in the experiment as the control group. They were recruited through the participant pool of the Department of Linguistics and Languages, SONA, and received course credit for their participation. All participants had a normal or corrected-to-normal vision and normal hearing. All participants provided informed consent. The study was cleared by the McMaster University Research Ethics Board.

Table 4.1 Summary of Persian-English bilinguals' language background information.

Variable	M (SD)	range	Variable	M(SD)	Range
Age of arrival	25.57(6.51)	12.42 − 42.75	AoA	10.92(2.63)	3.75 − 14.5
Length of stay	2.06(2.81)	0.08 − 12.75	Proficiency	4.53(0.64)	2.50 − 6
MoA:			Daily use		
Classroom	9.67(1.63)	6 −13	L1	0.54(0.22)	0.10 − 0.90
Naturalistic	2.11(2.79)	0.08 − 12.70	L2	0.45(0.22)	0.10 − 0.90
Mental use			Preference		
L1	5.50(1.33)	2 − 7	L1	4.47(1.09)	2.33 − 6.50
L2	3.52(1.29)	1.50 − 6.50	L2	4.64(1.21)	2.33 − 6.66

Proficiency (1: very low – 6: native-like), language use in all daily activities (0.00 – 1), mental language use (1: never – 7: always), and language preference (1: never – 7: always) were assessed based on self-rating reports by bilinguals. Manner of acquisition (MoA), formal classroom instruction or naturalistic exposure to L2 in an L2 environment, the age of L2 acquisition (AoA), the age of arrival in an L2 country, and the length of stay in an L2 environment were reported in years.

4.2.2 Memory stimuli and tasks

For the experimental target sentences, we created an original main list that included 160 unique sentences for each language (English and Persian), presented in 40 sequences of 4 instructions each, a total of 320 unique individual items. Then, based on the experimental temporal order condition (*matching* or *reverse*) and the type of connective (*before* or *after*), each original sequence yielded four variant sequences of instructions that included all conditions and connectives. Please see an example below:

2. Copy the diploma, then *BEFORE you print the certificate, sign the invoice*, and last, stamp the record.

2a. Copy the diploma, then *BEFORE you print the certificate, sign the invoice*, and last, stamp the record.

2b. Copy the diploma, then *print the certificate, BEFORE you sign the invoice*, and last, stamp the record.

2c. Copy the diploma, then *AFTER you print the certificate, sign the invoice*, and last, stamp the record.

2d. Copy the diploma, then *print the certificate, AFTER you sign the invoice*, and last, stamp the record.

We included each variant-sequence of sentences in a separate list by counterbalancing the temporal order conditions and the type of connectives. In this way, we created 4 lists of 40 sequences out of the original main list for each language. Each sequence of four instructions was composed of two linguistically simple sentences and a two-clause adverbial sentence expressing the temporal order of two actions. The experimental conditions and connectives were counterbalanced so that each list included only one instance of each original sequence. In this way, each list included 20 sequences in the *matching* condition and 20 sequences in the *reverse* condition, with equal numbers of *before* and *after* connectives (see Appendix F). Only one of the experimental conditions (matching vs. reverse order of linguistic phrases and the referred actions) and one of the connectives (before vs. after) were included in each sequence. Each subject received only one of the four lists of stimuli. Bilinguals received two distinct lists in each language with counterbalanced conditions in L1 and L2. They were never presented with the same sequence of instructions twice, either in L1 or L2.

The temporal order condition *reverse* vs. *matching*, and the type of connectives, *before* vs. *after*, created four types of complex sentences, i.e., *matching* with *before, reverse* with *before, matching* with *after,* and *reverse* with *after*. In the matching condition, the order of verbal mention of the events in complex sentences matched the intended temporal order of the actions. Thus, the order in which instructions were presented was consistent with the order in which they were performed. Conversely, in the reverse condition, the order of mention of the events did not match the temporal-order the actions had to be performed. In this condition, the action order had to be mentally re-arranged before acting out the sentences. Simple instructions signaled action order by their serial position in the sequence, whereas subordinate clauses beginning with *before* or *after* signaled which clause had to be acted on first in two-clause complex instructions. Each adverbial clause in the complex sentences described a distinct event. These were not causally or logically related to each other, had an arbitrary and unpredictable relation, could happen in any order and did not contain any references to antecedents across clauses. These characteristics were intended to reduce the bias for predicting and constructing mental representations based on world knowledge and/or causal relations. Therefore, participants were expected to have to rely on their linguistic knowledge of temporal order constructions in two-clause instructions, and the sequential position of simple instructions. The order of items in a given sequence within a list was fixed. However, we counterbalanced the order of conditions and also the assignment of stimulus lists for tasks and languages. Table 4.2 summarizes the conditions of the study along with some examples.

Table 4.2 The temporal order condition of the experiment (*matching* vs. *reverse*) and the type of adverbial connectives (*before* vs. *after*).

	Connective	
Condition	*before*	*after*
Matching	Copy the certificate, *before* you punch the invoice	*After* you copy the certificate, punch the invoice
Reverse	*Before* you punch the invoice, copy the certificate	Punch the invoice, *after* you copy the certificate

The complex instruction containing the temporal-order manipulation was embedded in a sequence of four instructions, i.e. between two simple instructions. In this way, we not only minimized the effects of primacy and recency on the recall accuracy of the complex instruction, but were also able to observe the effect of each condition on adjacent simple instructions.

The sequences of instructions in each list were related to daily or workplace activities, e.g., taking daily/regular medications or carrying out office-related duties. We used twelve action verbs (*print, copy, punch, sign, fax, scan, stamp, file, drink, use, administer*, and *take*), twenty-three type of documents (*lease, contract, claim, order form, complaint form, questionnaire, invoice, paycheque, pay stub, statement, declaration, licence, transcript, diploma, notice, checklist, proposal, agreement, record, resume, memo, manual*, and *schedule*), and six types of medications (*syrup, tablet, capsule, spray, nasal drops, ear drops,* and *eye drops*) to create the target sentences. We avoided including the same verb, document or medication twice in each sequence of instructions. Verbatim examples of two full instruction sequences are below.

3. *Sign the claim, <u>then after you copy the agreement, punch the report</u>, and last, fax the invoice.*
4. *Drink the syrup<u>, then before you administer the eye drops, use the spray</u>, and last, take the tablet.*

The English sentences were recorded by a female native speaker of Canadian English using the Audacity software with a sampling rate of 44100 Hz.

We inserted a 500 milliseconds (ms) silence after instructions 1, 2, and 3, and a 1000-ms silence after the last instruction in each sequence. A beep sound was added after the last silence to signal that the participant had to commence acting out the instructions. The voices were digitized and saved on the computer. Persian sentences were recorded by a female native speaker of Persian (Tehrani dialect). We followed the same recording and editing procedures as for the English stimuli. We used Microsoft PowerPoint software to run the enactment task. Each slide contained pictures of objects, documents, devices or medications, and the audio file of each sequence (see Appendix G). Overall, for each list, forty slides, each containing pictures of objects for sequences of 4 instructions, were created for the main task in each language in addition to experimental instruction slides and slides for practice trials.

4.2.3 Individual differences measures

We used the non-word repetition task of Experiment 1 to assess participants' phonological working memory. Also, we used the modified version of the English operation span task of Experiment 1 to assess participants' general working memory abilities. The Persian operation span task of Experiment 1 was used to measure bilinguals' working memory capacity in L1.

4.2.4 Design and procedure

The instruction-following experiment included three within-subjects variables: the language of instruction presentation (L1 Persian vs. L2 English), temporal order condition (matching vs. reverse), and type of connective (before vs. after), and one between-subjects variable: group (monolingual vs. bilingual). The procedure was similar to that of Experiments 1 and 2. Native speakers of English were tested in a single session in the Language, Memory, and Brain Lab at McMaster University. Persian-dominant bilinguals were assigned to perform the memory task first in L1 or L2 with the other language tested after a one-week interval. We used the language of testing to give the experimental instructions and to interact with the bilingual participants. The order of the language of presentation was counterbalanced between participants. All participants were naïve to the intent of the experiment and were told that they were going to perform a listening comprehension task. Upon arrival at the lab, the participants signed an informed consent form, filled the language background questionnaire (see Appendix B) and had time to ask questions. Then, they were seated in front of a 21.5" Macintosh computer screen and were asked to name all the objects, devices, documents, and medications, which would be presented with in the pictures in the test, to make sure that they were familiar with all their names. They were also instructed how to use the computer mouse to carry out actions that were referred to in the instructions, such as “signing”, “copying”, “stamping”, and “printing” documents and taking or administering medications. Each subject received a list of 40 sequences of four instructions each. These included two simple instructions and a two-clause instruction with a temporal order construction. The temporal order constructions were randomized with the restriction that *before/after* constructions of the same form were not allowed to follow each other consecutively. Before doing the main task, participants performed a practice trial. This consisted of two sequences of four instructions.

After reading the experimental instructions, the participant proceeded to the target task. They double-clicked on the sound icon in the top left corner of each PowerPoint slide and listened to the sequence of four instructions via headphones, with the goal to comprehend the instructions and remember their order in a sequence. They were told to listen to the taped set of instructions only once and to commence performing the actions on the screen after hearing a beep sound. This always followed the 1000 ms silence after the last instruction. Playback of the sound-file would automatically stop if the participant began to carry out the actions while instructions were being played. After the first 20 sequences, a message appeared on the screen asking if the participant wanted to take a short break;

otherwise, they could continue doing the task. Participants were reminded that they had to memorize the order of the events and perform the instructions in the correct serial order on the computer screen. The mouse movements were screen-recorded by Quick Time Player software to collect the responses.

The L1/L2 OSpan WM task followed the L1/L2 instruction performance task and was presented in the same language. We followed the same procedures as in Experiment 1 to conduct the OSpan task in English and Persian. The non-word repetition task was always conducted in the English session and with English-based stimuli only. After a practice trial, participants listened to the non-words one by one and immediately said them aloud. Participants' voices were recorded for data scoring and analyses.

4.2.5 Statistical considerations

The dependent variable was the recall accuracy in acting out the complex two-clause sentences containing temporal order in a given sequence of four instructions in each list of 160 items. As we were specifically interested in exploring the effects of the linguistic complexity of temporal order on recall accuracy in performing complex instructions, the analyses were mainly done on the data collected from the critical two-clause semantically complex sentences. We modelled the recall accuracy as the probability of correct response of each instruction in a given complex construction in each sequence, where a complex construction refers to a two-clause sentence containing the temporal order adverbials before and after. Contrary to most similar studies that have relied on averaging the performance of each subject, by employing ANOVA models for data analysis, we used the generalized linear mixed effects models with multiple variables and covariates. Given that the recall accuracy is bound by 0 and 1, violating the assumption of linear regression models and ANOVA (see Jaeger, (2008) for discussion), we used the generalized linear mixed effects regression models (Baayen, 2008; Baayen et al., 2008) as implemented in the *lme4* packag*e* (Bates et al., 2014) and R Core Team, (2017) with a binomial distribution to perform a linear mixed effects analysis of the relationship between independent variables and the recall accuracy in following linguistically complex instructions. This method of analysis allows exploring the effects of multiple factors and covariates while accounting for any variance contributed by both participants and items (in this case, each individual instruction). As random effects, we had intercepts for participants and items. As fixed effects in the models, we had the Temporal Order Condition (matching or reverse), the Type of connective (before or after), Language of presentation

(Persian or English), and Group (monolingual or bilingual), with scores on the complex working memory span, scores on phonological memory, and variables in the bilinguals' language background as continuous covariates. In addition, we used the serial Position of each instruction in a given sequence (1-4) and Trial (running number of an item in a list, 1-160) as control variables. Initially, we fitted each model with a maximal fixed effects structure and then removed the factors that did not significantly improve the model's performance. Where a model comparison was done, *p*-values were obtained by the likelihood ratio test of the full model with the effect in question against the model without the effect. Separate analyses were run to compare performance on English instructions between groups (monolinguals and bilinguals) and to compare within-group performance in L1 and L2 in the bilinguals.

The enactment of an individual instruction was scored correct if participants carried out the correct action on the mentioned object or document in its correct sequential order position. Thus, incorrect or missed instructions received zero points. Any errors in the sequential order of instructions, verbs, objects, documents or medications were considered as violations and received zero points. Also, participants received no credits if they disobeyed instructions and listened to the audio file of a sequence more than once. As a result of this scoring criterion, two bilinguals received zero points for some instructions. Participant number 23 received zero points for 19 items and participant number 26 received zeros for 25 items in a given list of 160 items. The probability of correct responses in each complex two-clause instruction for each participant was used in the data analysis.

The data collected from English and Persian operation span tasks were scored based on the number of words correctly recalled from memory in serial order. We followed the same scoring scheme in Experiments 1 and 2. We averaged the scores of the OSpan task in L1 for each participant, and used it as co-variates in the data analyses. The scores of the L2 OSpan task were also averaged for each bilingual participant and used in comparing L1 and L2 WM spans in bilinguals.

The data of the English non-word repetition task were scored based on the number of syllables accurately repeated in each non-word. We followed the same scoring scheme in Experiments 1 and 2. The probability of correct syllables in the non-word repetition task was averaged for each participant and used in the data analyses.

4.3 Results

4.3.1 The effects of linguistic complexity

Comparing recall accuracy of the simple instructions at the beginning and end of the instruction sequences with the complex middle instructions was not a research objective of this study. However, it can be noted that, as descriptive data in Table 4.3 shows, regardless of the temporal order condition (matching or reverse), the linguistic complexity appeared to influence participants' enactment accuracy, with lower recall accuracy in temporal order constructions. The effect of linguistic complexity appears to be present regardless of the language of presentation and group (see Table 4.3). Furthermore, bilinguals had higher recall accuracy in the L1 task than the L2 task.

Table 4.3 Mean proportion recall accuracy in performing sequences of four instructions containing the complex temporal order constructions in bilinguals (*N* = 35) and monolinguals (*N* = 42).

Group	**Language**	**Complex instructions**		**Simple instructions**		**Overall performance**	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Bilinguals	Persian	0.64	0.48	0.88	0.33	0.76	0.43
	English	0.52	0.50	0.82	0.38	0.67	0.47
Monolinguals	English	0.52	0.50	0.79	0.41	0.65	0.48

The critical question we asked in this study was whether the conceptual complexity of temporal order would affect remembering and performing two-clause complex instructions containing temporal order in L1 or L2. We tested the hypothesis that the reverse condition, where the surface order of event descriptions does not match the underlying order, is more demanding than the matching condition, where the surface order of event descriptions is congruent with the real-world event order. To this end, we fitted the data of semantically complex two-clause sentences into generalized mixed effects regression models. We included intercepts for participants and items as random effects and the fixed effects were the temporal order condition (matching or reverse) and the type of connective (*before* or *after*). We excluded the position and trial number in the models as all complex constructions were embedded as instructions 2 and 3 in a sequence of 4 instructions and we did not find the main effect of trial number on recall accuracy in experiments 1 and 2. Our other effect of interest was that of the language of presentation on recall accuracy in the enactment tasks in bilinguals. It was of particular interest whether the temporal complexity and language factors would interact with each other.

The results of the final models revealed a significant main effect of sequential temporal order condition (matching vs. reverse) on recall accuracy in following semantically complex instructions in bilinguals. As Table 4.4 and Figure 4.1 show, bilinguals were less accurate in acting out two-clause complex instructions in the reverse condition, where the conceptual order of events was incongruent with the order of mention of events. In other words, participants' recall was better when the factual order of events in the two-clause sentences matched the surface order of the events.

Table 4.4 Summary of the final logistic regression model of recall accuracy in enacting complex instructions in bilinguals (N = 35), reported as the regression coefficient estimates, standard errors, z values, and p values. The reference levels of fixed effects are marked.

Part A: random effects		
	Name	Std. Deviation
Item	Intercept	1.0734
Participant	Intercept	0.7877
	Language = *English*	0.9169

Part B: fixed effects	Estimate	Std. Error	z value	p value
Intercept	1.2173	0.1858	6.553	<.001
Language = *English*	-0.8161	0.2151	-3.793	<.001
Connective = *before*	-0.2651	0.1480	-1.791	0.0733
Temporal order = *reverse*	-0.6655	0.1729	-3.849	<.001
Language *English* * temporal order *reverse*	0.2749	0.2040	1.347	0.1779
Connective *before* * temporal order *reverse*	0.1630	0.2032	0.802	0.4226

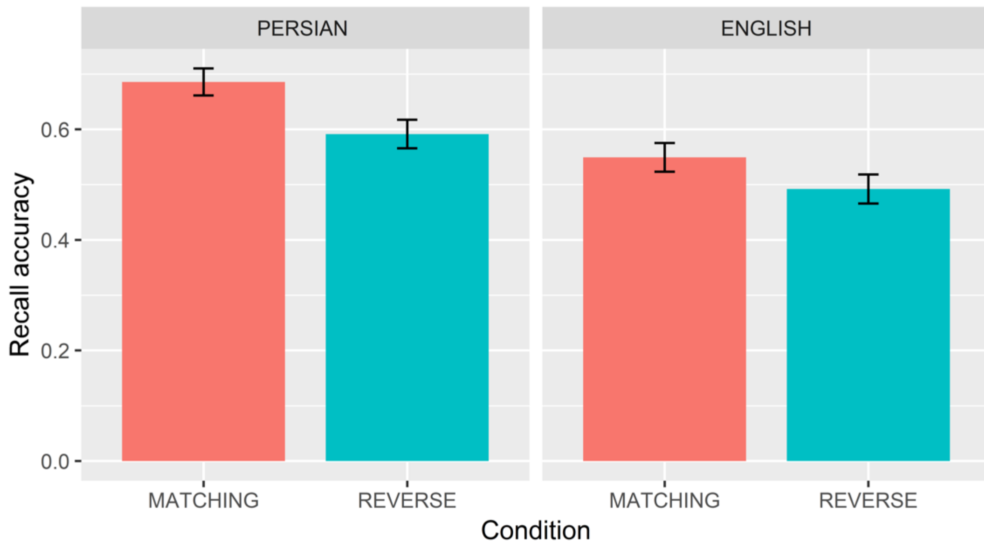

Figure 4.1 The effect of the condition and the language of presentation on recall accuracy in acting out complex instructions in Persian-English bilinguals.

As Table 4.4 and Figure 4.1 show, the language of presentation also significantly influenced the response accuracy in bilinguals. Participants remembered and performed fewer complex instructions when the target sentences were presented in their non-dominant language. However, the interaction between the language of task and the temporal order condition was not significant, indicating that the effect of temporal order was not reliably bigger in the L2 condition. In fact, this difference showed an unexpected trend to be bigger in the L1 comparison.

We further examined if the temporal connective (*before* vs. *after*) would influence recall accuracy of complex temporal-order clauses in the bilinguals' data. The results of the final model showed that the main effect of the connective on recall accuracy was marginal (p = .07), and the data suggest that *before* may lead to poorer performance, although this was not strongly shown. The interaction between the temporal order condition and the type of connective did not reach significance either, indicating that the reverse condition was demanding in complex constructions with both of *before* and *after* connectives (see Table 4.4).

In a separate model, we investigated the effect of conceptual complexity on the enactment accuracy in complex instructions in English monolinguals. The results of the models indicated that the recall accuracy in remembering and acting out the two-clause complex instructions was significantly affected by the temporal order condition (see Table 4.5). Participants were less accurate when the conceptual order of the events in the two-clauses was incongruent with the order of mention of the events, in the reverse condition. We again found no reliable (p = .0815) main effect of the type of connective on the recall accuracy in following complex instructions, and this time performance was somewhat better with *before*. The connective did not significantly interact with temporal order condition, suggesting that, regardless of the type of the connective, the instructions in the reverse condition were more difficult to remember and perform than those in the matching condition.

Table 4.5 Summary of the final logistic regression model of recall accuracy in enacting complex instructions in monolinguals (*N* = 42), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference levels of the fixed effects are marked.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	1.207
Participant	Intercept	0.720

Part B: fixed effects

	Estimate	Std. Error	*z* value	*p* value
Intercept	0.2075	0.1937	1.071	0.2842
Connective = *before*	0.3905	0.2241	1.742	0.0815
Temporal order condition = *reverse*	-0.6005	0.2246	-2.674	0.0075
Connective *before** temporal order condition *reverse*	0.0290	0.3167	0.092	0.9270

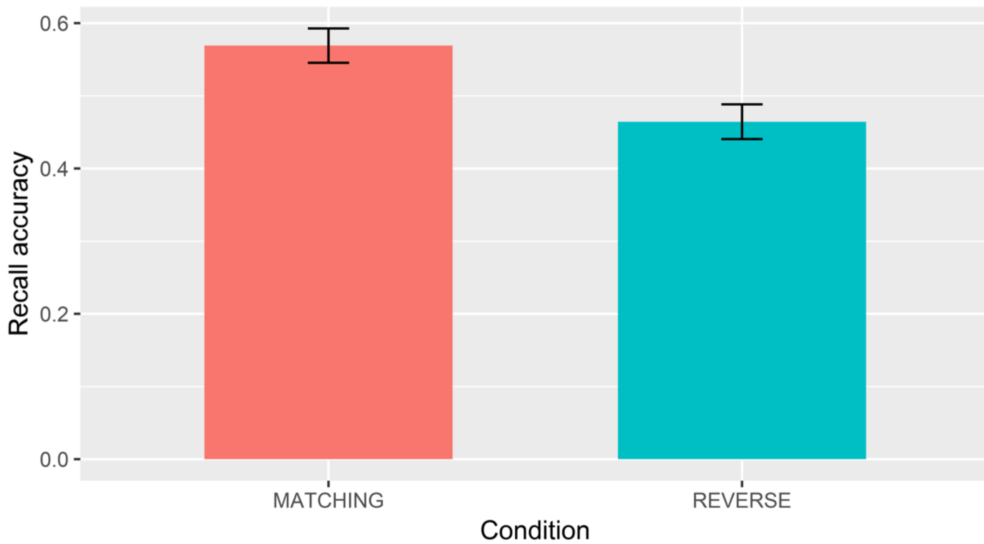

Figure 4.2 The effect of the temporal order condition on recall accuracy in acting out complex English instructions in English native speakers.

A separate analysis was conducted to compare the performance of bilinguals and monolinguals in only English instructions. We fitted the English data for complex instructions into generalized mixed effects models with participants and items as random effects, and group, temporal order condition, and connective as the fixed effects. As Table 4.6 shows, the main effect of group, as a between-subjects factor, on recall accuracy in following complex English instructions did not reach a significance level (also see descriptive data in Table 4.3). Recall accuracy was affected by the sequential temporal order condition with the instructions in the reverse order being more difficult to remember and perform. The main effect of connective did not reach significance. However, there was a significant interaction between the type of connective and group, showing that English monolinguals tended to be more accurate in recalling constructions including the *before* connective whereas the trend was the opposite for the English L2 speakers. There was also a significant interaction between the temporal order condition and group, with a lower recall accuracy in the reverse order condition in the monolingual group (see Figure 4.3).

Table 4.6 Summary of the final logistic regression model of recall accuracy in performing complex English instructions in monolinguals (*N* = 42) and bilinguals (*N* = 35), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level of categorical variables is marked.

Part A: random effects		
	Name	Std. Deviation
Item	Intercept	1.0879
Participant	Intercept	0.8223

Part B: fixed effects	Estimate	Std. Error	*z* value	*p* value
Intercept	0.4406	0.2032	2.168	0.0301
Connective = *before*	-0.3031	0.2031	-1.492	0.1356
Temporal order condition = *reverse*	-0.4505	0.2034	-2.215	0.0267
Group = monolingual	-0.1741	0.2152	-0.809	0.4184
Connective *before** temporal order condition *reverse*	0.2750	0.2716	1.013	0.3113
Temporal order condition *reverse** group *monolingual*	-0.2423	0.1196	-2.026	0.0428
Connective *before** group *monolingual*	0.5508	0.1196	4.607	<.001

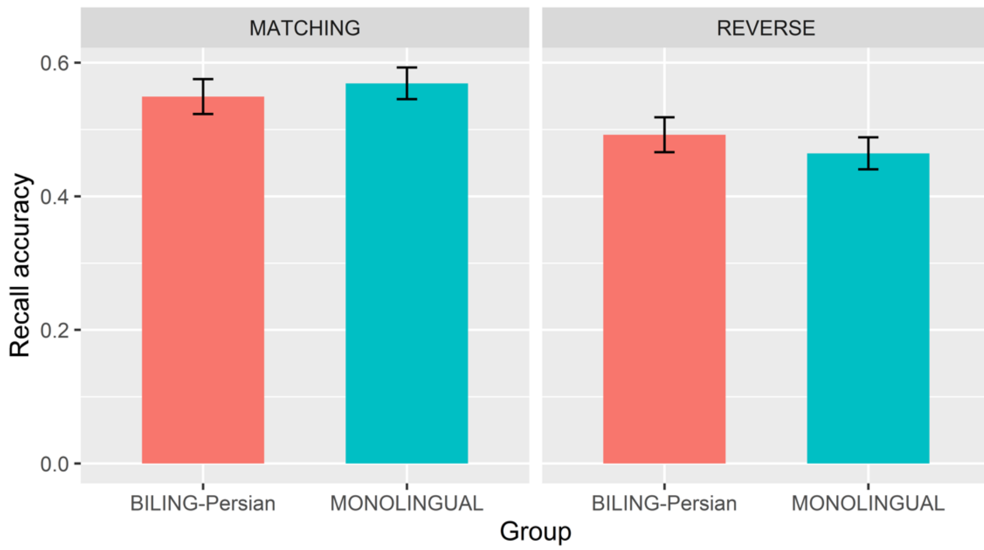

Figure 4.3 The interaction between the group and temporal order condition in English data.

In sum, the results of the models showed that the linguistic complexity of expressions of temporal order significantly affected accuracy in recalling and following semantically complex instructions. Participants remembered and executed fewer complex instructions when the underlying order of events was incongruent with the surface order of event mentions in the reverse condition. Thus, the reverse constructions seemed to be more demanding for the cognitive system. The effect of the temporal order condition was present regardless of the type of temporal order connectives *before* and *after* and the language of presentation for bilinguals. In bilinguals, the language of presentation impacted recall accuracy resulting in lower enactment accuracy for complex instructions in the non-dominant compared to the dominant language. Temporal order condition interacted with the group in the only English tasks. The English monolingual group had lower recall accuracy in reverse constructions than the bilingual group. The significant interaction between the group and the type of connective in only English tasks revealed that English monolinguals had higher recall accuracy in two-clause constructions containing the subordinate conjunction '*before'*. This suggests that understanding *before*-constructions was more difficult for the bilingual group.

4.3.2 L2 background variables

We explored the relationship between individual differences in L2 acquisition history and bilinguals' following of complex two-clause instructions in a non-dominant language. Bilinguals' recall accuracy in acting out complex two-clause temporal order constructions was fitted into generalized mixed effects regression models with bilinguals' L2 language acquisition variables as predictors. We treated participants and items as random effects. The predictors were participants' current age, education (undergraduate, Masters, and Ph.D.), gender, the age of arrival in an English-speaking country (Canada), length of residence in Canada and/or an

English-speaking country (in years), manner of acquiring English as a second language (classroom instruction or natural exposure (in years)), mean onset of L2 age of acquisition, mean degree of proficiency in listening, speaking, reading, and writing skills (based on 1 – 6 self-reported rating), L1 and L2 daily use (based on self-reported percentage), mental L1 and L2 use (based on 1 – 7 self-reported rating), and L1 and L2 preference in academic and non-academic situations (based on 1 – 7 self-reported rating) (see Table 4.1). An exploratory analysis of L2 background variables showed that there was a correlation between the L2 AoA and L2 proficiency, r = -.40 [95% CI: -0.65 – -0.07], suggesting that bilinguals who acquired English earlier in life rated themselves more proficient than those who acquired their L2 later in life. However, the correlation between L2 proficiency and the manner of L2 acquisition was significantly different, r = .39 [95% CI: 0.07 – 0.64] for the naturalistic exposure and r = -.02 [95% CI: -0.35 – 0.32] for the classroom instruction, suggesting that bilinguals who acquired English as an L2 in a more naturalistic way considered themselves more proficient with those who just learned the L2 by formal classroom instruction.

The results of the models showed that some factors, such as mean L2 age of acquisition (AoA) and L2 proficiency in four skills (listening, speaking, reading, and writing, daily L1 and L2 use, the age of arrival in an L2 environment (e.g., Canada), and the time bilinguals spent learning L2, were correlated with recall accuracy for performing L2 complex instructions containing temporal order constructions.

The results of the models showed that the L2 AoA significantly affected recall accuracy in following complex temporal order instructions (see Table. 4. 7). As Figure 4. 4 shows, bilinguals who acquired English as an additional language earlier in life had better response accuracy than those who started learning English later in life. However, there was not a significant interaction between the L2 AoA and the temporal order condition, $b = -0.0287$, SE = 0.0358, $z = -0.803$, $p = 0.4221$ (model not included), failing to detect that L2 AoA would have influenced the recall accuracy differently in matching and reverse conditions. The likelihood ratio test showed that there was not a significant difference between the simple model and the model including the interaction, $X^2 (1) = 0.6275$, $p = 0.4283$, therefore, we report the results of the simple model in Table 4.7.

Table 4.7 Summary of the final logistic regression model of instruction recall accuracy and L2 AoA in bilinguals (*N* = 35), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level of the categorical fixed effect is marked.

Part A: random effects				
	Name		Std. Deviation	
Item	Intercept		0.9406	
Participant	Intercept		0.8061	

Part B: fixed effects	Estimate	Std. Error	*z* value	*p* value
Intercept	2.0187	0.6331	3.189	0.0014
Temporal order condition = *reverse*	-0.3014	0.1369	-2.202	0.0277
L2 AoA	-0.1612	0.0559	-2.884	0.0039

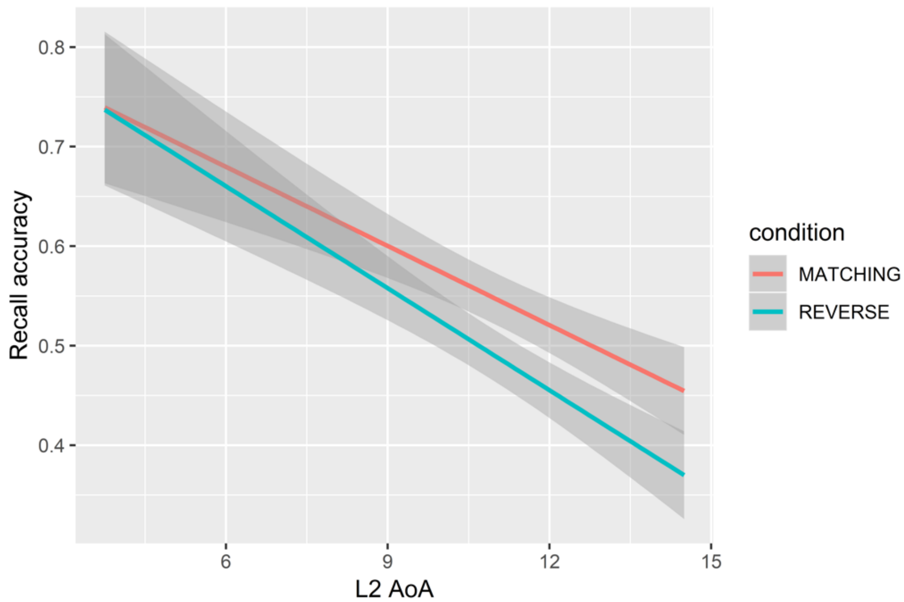

Figure 4.4 The relationship between L2 AoA and recall accuracy in acting out L2 complex instructions in Persian-English bilinguals.

Another factor that significantly influenced the enactment accuracy for L2 complex instructions was bilinguals' functional language skills. As Table 4.8 and Figure 4.5 show, the degree of L2 proficiency significantly correlated with the recall accuracy in performing complex temporal order constructions in L2. Bilingual participants with higher L2 skills recalled and acted out more complex instructions than less-skilled participants. However, the interaction between L2 proficiency and temporal order condition did not reach significant level, b = -0.1573, SE = 0.1505, z = -1.045, p = 0.2960 (model not shown), compatible with the impression that L2 proficiency influenced the recall accuracy in both matching and reverse conditions. The likelihood ratio test showed that there was not a significant difference between the simple model with only main effects and a complex model including the interaction between the factors, $X^2\,(1) = 1.0568$, $p = 0.304$. We report the results of the final simple model in Table 4.8.

Table 4.8 Summary of the final logistic regression model of instruction recall accuracy and L2 proficiency in bilinguals ($N = 35$), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level of the categorical fixed effect is marked.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.9408
Participant	Intercept	0.7721

Part B: fixed effects

	Estimate	Std. Error	*z* value	*p* value
Intercept	-3.2082	1.0204	-3.144	0.0017
Temporal order condition = *reverse*	-0.3020	0.1369	-2.206	0.0274
L2 proficiency	0.7649	0.2224	3.440	<.001

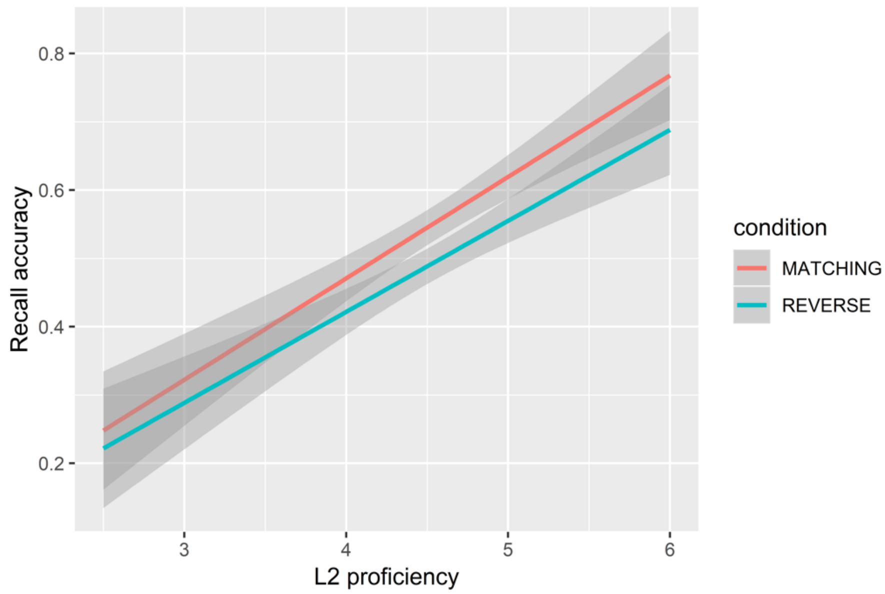

Figure 4.5 The relationship between the degree of L2 proficiency and recall accuracy in following complex verbal instructions in the bilinguals' L2.

Daily L2 use also influenced the accuracy of acting out complex instructions (see Tables 4.9). As Figure 4.6 shows, bilinguals who used English more frequently in daily and academic activities were more accurate in performing complex instructions. There was a marginal interaction effect between temporal order condition and daily L2 use pointing to a bigger advantage for the 'matching' order condition for bilinguals who were using their non-dominant language more frequently. Although the interaction between daily L2 use and temporal order condition did not reach significance, we expected more skilled bilinguals to have a better performance in the reverse condition. However, it seems that more proficient bilinguals had the advantage to follow less difficult matching condition.

Table 4.9 Summary of the final logistic regression model of recall accuracy and daily L2 use in bilinguals (*N* = 35), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference levels of categorical fixed effects are marked.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.9460
Participant	Intercept	0.8402

Part B: fixed effects

	Estimate	Std. Error	*z* value	*p* value
Intercept	-0.6154	0.3681	-1.672	0.0945
Temporal order condition = *reverse*	0.0025	0.2352	0.011	0.9915
Daily L2 use	1.9359	0.7212	2.684	0.0073
Temporal order condition *reverse** daily L2 use	-0.6695	0.4214	-1.589	0.1122

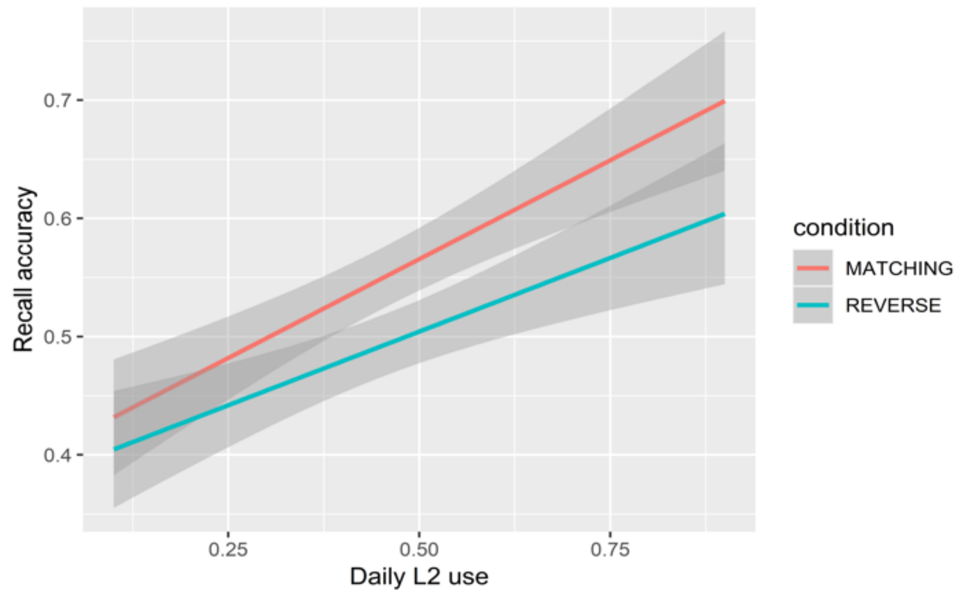

Figure 4.6 The relationship between daily L2 use and following complex verbal instructions in bilinguals' non-native language.

The overall time bilinguals had spent learning English as a foreign or second language impacted the response accuracy in performing L2 instructions, $b = 0.1250$, SE = 0.0627, $z = 1.995$, $p = 0.0461$(model not shown). Bilinguals who had been learning English (either through classroom instruction or naturalistic exposure) recalled instructions better than those who had been less exposed to the L2 (see Figure 4.7). However, the interaction between the number of years participants had learned English and temporal order condition did not reach a significant level, and that the effect could not be reliably detected.

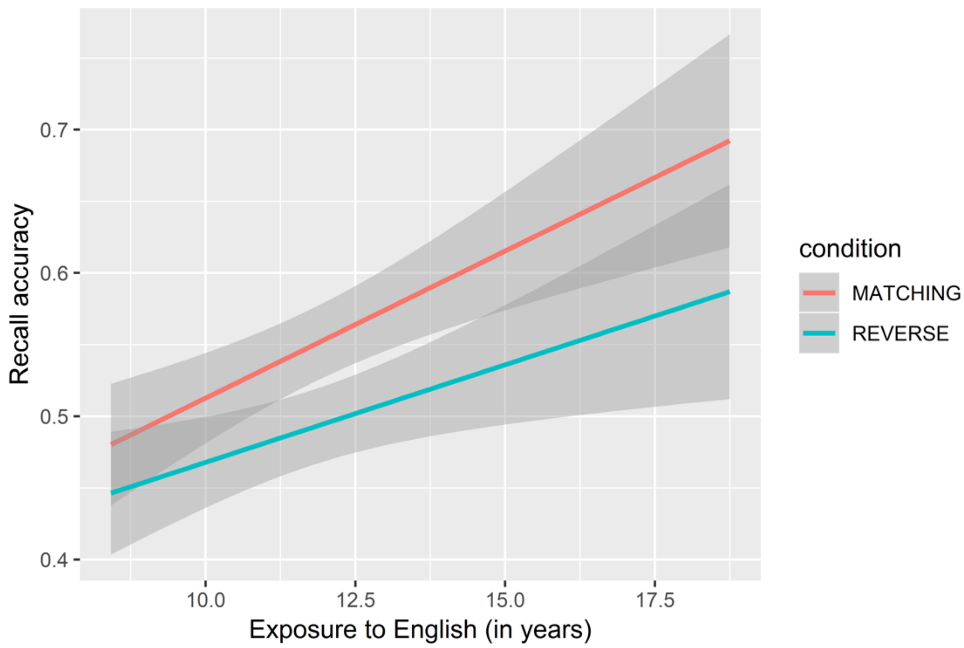

Figure 4.7 The relationship between exposure to L2 and following complex verbal instructions in bilinguals' non-dominant language.

Another factor that influenced the recall accuracy in performing L2 complex instructions was the age of arrival in an L2 environment (Canada). As Table 4.10 and Figure 4.8 show, bilinguals who arrived in an English-speaking country later in life acted out fewer correct complex instructions than participants who arrived earlier in life. The age of arrival significantly interacted with the temporal order condition. Bilinguals who had arrived in Canada earlier in life had a greater response accuracy advantage in the matching condition. The likelihood ratio test revealed that the model including the interaction factor was indeed a better fit than the simpler main-effect model, X^2 (1) = 6.3851, *p* =0.0115.

Table 4.10 Summary of the final logistic regression model of recall accuracy and the age of arrival in bilinguals (*N* = 35), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference levels of categorical fixed effects are marked.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.9478
Participant	Intercept	0.7846

Part B: fixed effects

	Estimate	Std. Error	*z* value	*p* value
Intercept	2.5816	0.6181	4.176	<.001
Temporal order condition = *reverse*	-1.2366	0.3921	-3.153	0.0016
Age of arrival	-0.0907	0.0233	-3.900	<.001
Temporal order condition *reverse** age of arrival	0.0365	0.0143	2.550	0.0108

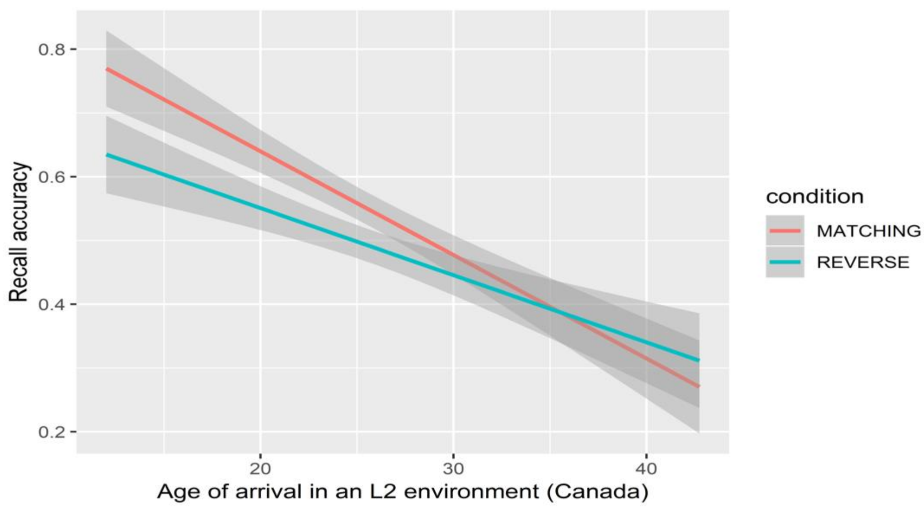

Figure 4.8 The relationship between the age of arrival in Canada and acting out L2 complex temporal order instructions in bilinguals.

We found no reliable or significant effects of other factors in bilinguals' language background on the recall accuracy in performing L2 two-clause instructions containing temporal order connectives. In addition, these elements did not significantly interact with the temporal order of event mentions in two-clause sentences relative to the order of the referenced events in the real world.

In brief, it seems that more variables in the bilinguals' language history correlated with following conceptually complex instructions in this study compared to studies in Experiments 1 and 2. Remember that we found no significant relationship between the recall accuracy and the elements in the bilinguals' L2 background in Experiment 1, in which the bilinguals were advanced Persian learners of English like in the present experiment. The significant effects of variables such as the L2 AoA, proficiency, language use, exposure, and age of arrival on the response accuracy suggest that processing linguistically complex constructions and/or performing more complex tasks in L2 rely on more language skills.

4.3.3 The effect of working memory capacity: The Operation Span Task

One of the research questions of this study was to what extent individual differences in working memory capacity would influence following linguistically complex instructions containing temporal order. To this aim, we examined the relationship between working memory capacity and following two-clause oral instructions in bilinguals and monolinguals. We hypothesized that participants with a larger working memory capacity would exhibit better recall accuracy in acting out conceptually complex oral instructions. We further investigated if the language of the WM span task (operation span) would influence the WM score, by assessing WM separately in L1 and L2.

As descriptive data in Table 4.11 show, the language of testing affected bilinguals' WM span. Bilinguals had higher WM span scores in the Persian complex span task than the English complex span task.

4.11 Summary of the complex span tasks in bilinguals (*N* = 35) and monolinguals (*N* = 42), reported as mean WM span (standard deviation), arithmetic accuracy, and mean RT for equations and responses in milliseconds

Group	**Language**	**WM span (SD)**	**Arithmetic accuracy (range)**	**Mean RT (ms)**
Bilinguals	Persian	0.70 (0.12)	0.93 (0.63-1)	6831
	English	0.59 (0.12)	0.93 (0.63-1)	8550
Monolinguals	English	0.61 (0.14)	0.82 (0.42-1)	7429

A one-way within-subjects analysis of variance (ANOVA) showed that the effect of the language of presentation on WM span tasks was significant, F (1, 68) = 11.96, p <.001. Persian L1 bilinguals recalled fewer words for memory in the correct serial order in English trials than in Persian trials (see Figure 4.9). Despite the differences in the scores of L1 and L2 complex span tasks, L1 and L2 WM spans were highly correlated, r = .59 [95% CI: 0.33 – 0.77].

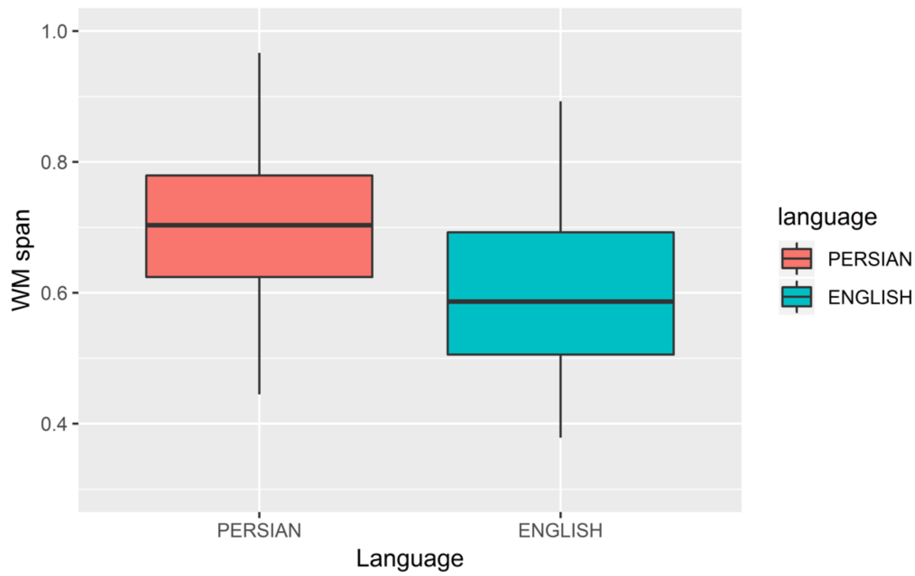

Figure 4.9 The effect of the language of task on bilinguals' WM span.

However, a one-way between-groups ANOVA, comparing bilinguals and monolinguals in the English complex span task, indicated that the effect of the group on WM span was not significant, and there were no significant differences between the means of English WM span in the bilingual and monolingual groups, $F(1, 75) = 0.349$, $p = 0.557$.

In this study, we further examined the hypothesis that individual differences in working memory capacity would affect the recall accuracy in following temporal order instructions with an advantage for individuals with a greater WM span. To this end, we fitted the L1 and L2 data of the bilingual group in the instruction-following tasks with the scores of L1 WM span as a predictor into generalized mixed effects regression models. We had participants and items as the random effects and the scores of the L1 OSpan task predicting recall accuracy in remembering and acting out complex instructions. The results of the bilingual models showed that the main effect of WM span on the recall accuracy in performing complex instructions was not significant (see Table 4.12). We did not find a significant interaction between the WM span and the language of task. The interaction between the WM span and the temporal order condition was non-significant as well. Thus, there was no reliable effect of L1 WM capacity, as reflected in OSpan, on recall of L2 instructions with complex temporal structures.

Table 4.12 Summary of the final logistic regression model of recall accuracy and WM span in bilinguals (*N* = 35), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level for categorical fixed effects is marked.

Part A: random effects		
	Name	Std. Deviation
Item	Intercept	1.0813
Participant	Intercept	0.9993
	Language = *Persian*	1.0819

Part B: fixed effects	Estimate	Std. Error	*z* value	*p* value
Intercept	-0.1349	0.7506	-0.180	0.857
Temporal order condition = *reverse*	-0.5771	0.3938	-1.465	0.143
WM span	1.2901	1.0643	1.212	0.225
Temporal order condition *reverse** WM span	0.1751	0.5468	0.320	0.749

In a separate model, we explored the relationship between recall accuracy and L1 WM span in English monolinguals. The results of the model revealed that in this analysis, there was a significant main effect of temporal order condition, showing a disadvantage for the reverse temporal order condition. There was a significant relationship between WM span and recall accuracy in performing complex instructions (see Table 4.13 and Figure 4.10). Participants with a greater

WM span were more accurate in executing complex instructions. However, we again found no significant interaction between WM span and the temporal order condition. The comparison of models showed no significant differences between the simpler models and the complex model including an interaction factor, $X^2(1) = 0.3871$, $p = 0.5338$.

Table 4.13 Summary of information of final logistic regression model of recall accuracy and WM span in monolinguals ($N = 42$), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level for categorical fixed effects is marked.

Part A: random effects		
	Name	Std. Deviation
Item	Intercept	1.2266
Participant	Intercept	0.6008

Part B: fixed effects	Estimate	Std. Error	*z* value	*p* value
Intercept	-1.1836	0.4994	-2.370	0.0178
Temporal order condition = *reverse*	-0.8178	0.3986	-2.051	0.0402
WM span	2.5921	0.5997	0.632	<.001
Temporal order condition *reverse** WM span	0.3792	0.5997	0.632	0.5272

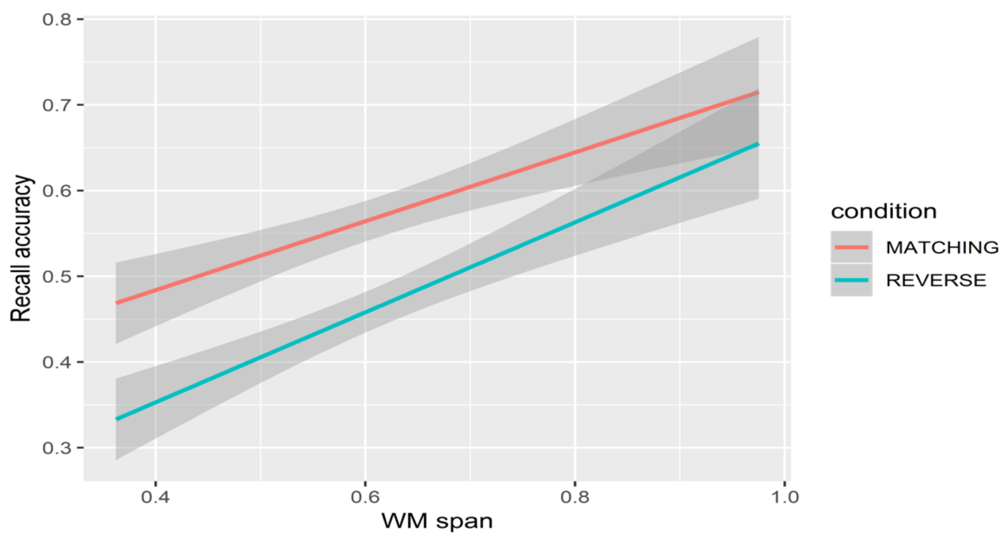

Figure 4.10 The relationship between WM span and recall accuracy in acting out complex temporal order instructions in English monolinguals.

The comparison of L1 and L2 WM span in bilinguals revealed that the language of presentation influenced WM span scores. Bilinguals exhibited lower WM span scores when the WM task was presented in the non-dominant language. Overall, the results of the models on the relationship between WM span and response accuracy in complex instruction recall were mixed. While we found no significant main effect of L1 WM span on recall accuracy in the bilingual group, with apparently better L1 WM scores, the correlation between individual differences in English WM span and response accuracy to English instructions was significant in the monolingual group, with apparently poorer L1 WM scores. It seems that individuals with a greater WM span were able to act out more complex oral instructions in the group with relatively lower L1 WM scores. However, WM span did not significantly correlate with temporal order condition in following complex instructions in this group either. Since the bilingual group generally had good L1 WM scores, there would have been some range compression that made it less likely to find correlations between WM span and temporal order condition constructions.

4.3.4 The influence of phonological WM: The Non-word Repetition Task

To compare the phonological memory scores in bilinguals and monolinguals' non-word repetition task, we conducted a one-way between-groups analysis of variance (ANOVA). The model showed a significant main effect of the group, $F(1, 75) = 56.3$, $p < .001$. As Figure 4.11 shows, bilinguals were significantly less accurate in repeating English-based pseudo-words in the non-word repetition task than their monolingual peers.

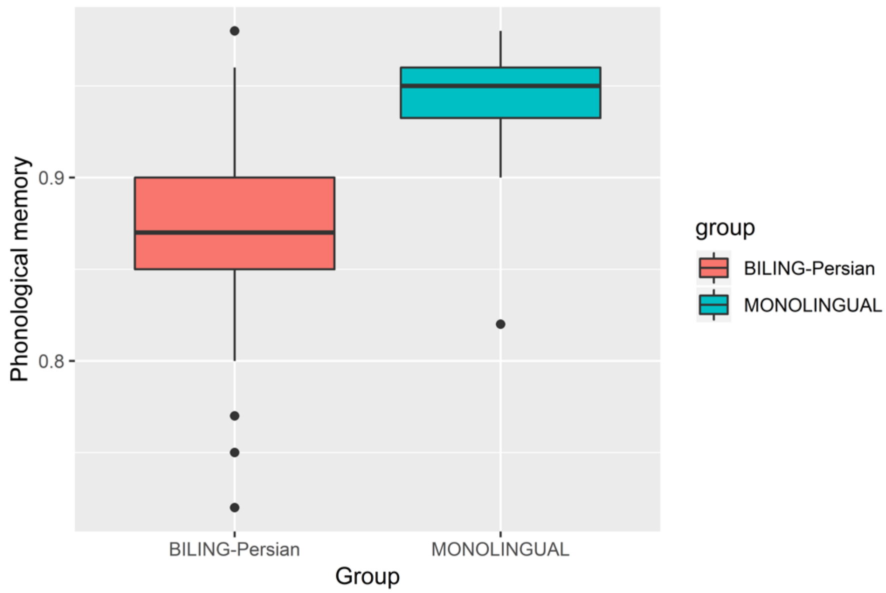

Figure 4.11 Comparison of phonological memory scores in Persian-English bilinguals and monolinguals in a repetition test with pseudowords following English phonotactics.

To investigate the effect of phonological memory on the recall accuracy of complex temporal order clauses, we fitted the English data of the bilingual group into generalized mixed effects regression models with participants and items as random effects and the scores of the non-word repetition task as a predictor of the recall accuracy for English instructions. The results of the models revealed a significant main effect of phonological memory on L2 enactment accuracy (see Table 4.14). As Figure 4.12 shows, participants with higher phonological memory scores performed better in following complex instructions. However, as Table 4.14

shows, there was a marginal interaction between phonological memory and condition. Phonological memory performance seems to have a stronger relationship with recall accuracy for matching than for reverse stimuli. Thus, memory for verbal/surface representations may have directly helped recall in the matching trials, but less when the order of events had to be re-arranged in the reverse constructions.

Table 4.14 Summary of the final logistic regression model of L2 recall accuracy and phonological memory in bilinguals (N = 35), reported as the regression coefficient estimates, standard errors, z values, and p values. The reference level for categorical variables is marked.

Part A: random effects

	Name	Std. Deviation
Item	Intercept	0.9442
Participant	Intercept	0.8631

Part B: fixed effects

	Estimate	Std. Error	z value	p value
Intercept	-5.533	2.530	-2.187	0.0287
Temporal order condition = *reverse*	2.391	1.502	1.592	0.1114
Phonological memory	6.662	2.903	2.295	0.0218
Temporal order condition *reverse** phonological memory	-3.095	1.720	-1.800	0.0719

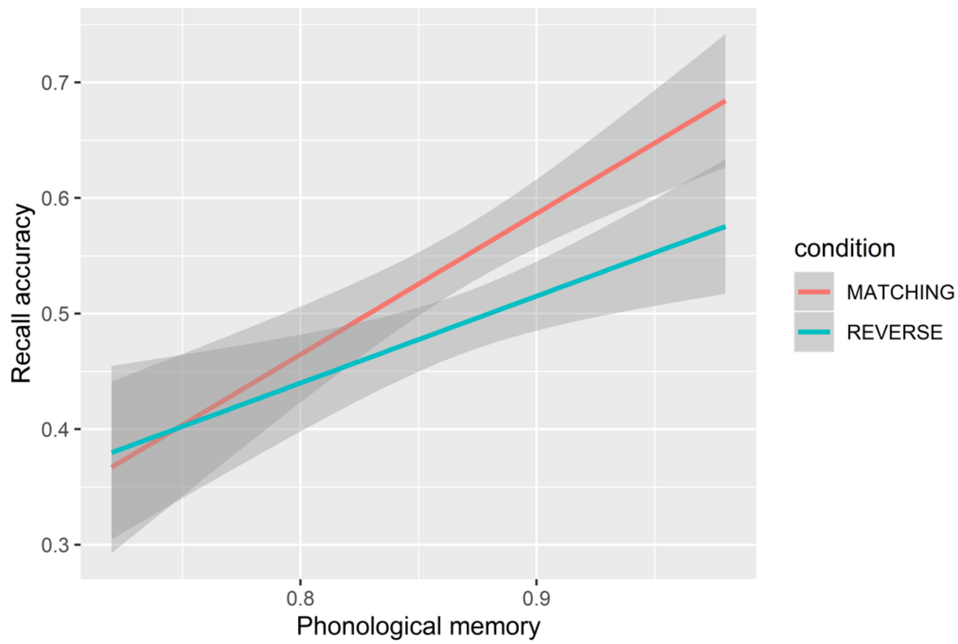

Figure 4.12 The relationship between phonological memory and recall accuracy in following L2 complex instructions in Persian-English bilinguals.

To further examine the effect of phonological memory on instruction recall accuracy of complex temporal order constructions in only English tasks, we fitted the English data of both groups into generalized mixed effects regression models with participants and items as random effects and the scores of the non-word repetition task as a predictor of recall accuracy for English instructions. The results of the models revealed a significant main effect of phonological memory on enactment accuracy (see Table 4.15). As Figure 4.13 shows, participants with higher phonological memory scores had better recall accuracy in following complex instructions. As Table 4.15 shows, there was also an interaction between phonological memory and condition. Phonological memory performance seems to have a stronger relationship with recall accuracy for matching than for reverse stimuli. The effect of the group as a between-subjects factor did not reach significance. Also, the interaction between the group and phonological memory did not reach significance.

Table 4.15 Summary of the final logistic regression model of the recall accuracy and phonological memory in bilinguals (*N* = 35) and English monolinguals (*N* = 42), reported as the regression coefficient estimates, standard errors, *z* values, and *p* values. The reference level for categorical variables is marked.

Part A: random effects				
	Name	Std. Deviation		
Item	Intercept	1.0917		
Participant	Intercept	0.7961		
Part B: fixed effects				
	Estimate	Std. Error	*z* value	*p* value
Intercept	-5.8417	2.2029	-2.652	0.0080
Temporal order condition = *reverse*	2.3625	0.9711	2.433	0.0150
Phonological memory	7.0552	2.5219	2.796	0.0052
Group = *monolinguals*	5.8543	4.0839	1.433	0.1517
Temporal order condition *reverse** phonological memory	-3.0801	1.0558	-2.917	0.0035
Group *monolinguals** *phonological memory*	-6.6530	4.4046	-1.510	0.1309

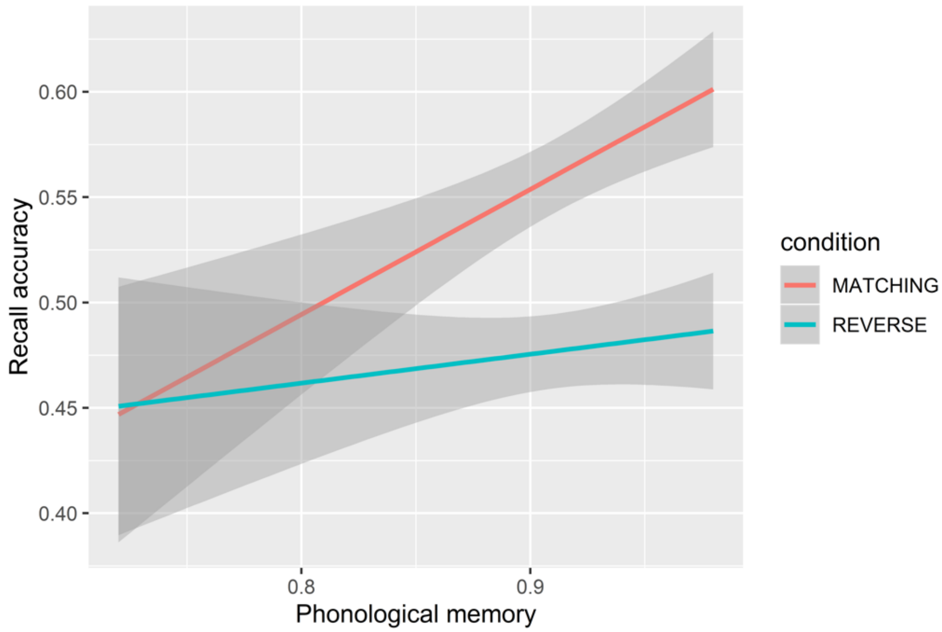

Figure 4.13 The relationship between phonological memory and recall accuracy in following English complex instructions in the combined group of Persian-English bilinguals and English monolinguals.

4.4 Discussion

The current study is one of the first empirical studies investigating the effect of linguistic complexity on understanding and executing complex instructions in a second language. More specifically, we manipulated the memory for order of event mention in matching and reverse conditions, as the order of the events in two-clause temporal order constructions could be either congruent or incongruent with the real-world conceptual order. We further explored the impact of language of testing and variables in skilled Persian-English bilinguals' linguistic background on the accuracy of comprehending and executing sequences of multi-step complex oral instructions. In addition, we wanted to know if there was a relationship between individual scores in functional WM or phonological short-term memory and recall accuracy in performing two-clause complex instructions in bilinguals and monolinguals. Presenting the instruction sentences verbally and requiring the participants to immediately execute them made the tasks more challenging. As the recall accuracy data for each instruction was binomial, we employed generalized mixed effects logistic regression models to explore the relationship between the independent and dependent variables. This allowed us to add statistical power as compared with similar previous studies.

A central question was whether having to process a second language would harm memory for complex instructions. We found that the bilinguals consistently had better recall accuracy in their dominant language. Thus, the language of testing (L1 vs. L2) significantly influenced the recall accuracy for conceptually complex instructions. Bilinguals' performance was inferior when the sentences were presented in their L2. The effect of the language of presentation was also evident in the complex span task, reflected in higher working memory span scores in L1. The results are consistent with findings from studies that showed that input in a well-mastered but non-dominant language puts a strain on the cognitive system (Roberts, 2012) and depletes working memory resources. (Paap, Sawi, Dalibar, Darrow, & Johnson, 2015; Service, Simola, Metsänheimo, & Maury, 2002). The effect of the language of testing on bilingual processing was further reflected in the finding that L2 processing was slower and less automated than L1 processing, especially in less proficient L2 learners. Thus, the integration of various information types appeared to be more costly, slower and more conscious in L2 than in L1 sentence processing (Fernández, 2003). It could be speculated that slower processing in L2 might occupy a bottleneck in central executive capacity for longer than the faster L1 processes, leaving more capacity for maintenance of instructions. The fact that language and temporal order difficulty did not interact, suggests that these two factors affected different stages of processing. For example, language processing may have been done before construction of mental representations began, so the two would not have competed for WM resources.

Our findings were consistent with the findings of Mandler's (1986) reading experiments in which the reading comprehension was faster when the order of mention of events was congruent rather than incongruent with the factual order of events. The results of our analyses also indicated that reversed temporal order of event descriptions in two-clause sentences significantly impaired remembering. This was seen both in the ability to maintain the verbatim details of what was said, and the correct ordering of the events at recall. Instruction enactment was poorer in both bilingual and monolingual groups. A recall advantage for the matching condition relative to the reverse condition was consistently observed for all participants, regardless of group (bilinguals or monolinguals) and the language of the task in the bilingual group (the interaction effects with both the group and language were non-significant).

The pronounced difficulties in following spoken instructions in the reverse condition can be associated with a revision process. When the order of the temporal clauses does not match the factual/conceptual order, the event representations have to be mentally re-arranged or revised. Revising the order of the events to construct a coherent mental representation can be expected to put an extra burden on the cognitive system. Hence, more working memory resources are required to maintain the critical information and the order of events in the main and subordinate clauses, retrieve the information and order, and ultimately re-arrange the order of the events in two-clause commands before commencing the execution. Any difficulties in maintaining and/or retrieving the critical information and failure in re-arranging the order of the events in the reverse condition would result in a poor or fragile mental representation, which would be reflected in poor execution of instructions.

The difficulty in following semantically complex instructions with reversed event order was further confirmed by the relationship between individual differences in working memory capacity and ability to follow the instructions. Complex span tasks, such as the reading span task or the operation span task, force WM storage operations in the face of processing (or distraction) to engage executive attention processes (Conway et al., 2005), thus, making the task more demanding for the cognitive system. Our results are consistent with the predictions of computational models, such as Just and Carpenter's (1992) sentence comprehension model, and the approach that considers the role of WM in complex cognitive tasks from the perspective of individual differences as measured by complex span tasks (Daneman & Carpenter, 1980; Engle, Carullo, & Collins, 1991; Miyake & Shah, 1999). According to this approach, the demands in complex span tasks, like the operation span task in this study, resemble the WM demands during the performance of complex cognitive tasks that also impose simultaneous demands on both processing and storage. Research has found that there is relationship between working memory capacity and language processing, as individuals with a higher memory span are also more accurate in sentence comprehension, resolving syntactic ambiguity, integrating pragmatic, lexical-semantic, and syntactic information to accomplish efficient processing, and show more sensitivity to syntactic violations (Dai, 2015; Dussias & Piñar, 2010; Farmer et al., 2017; Foroughi, Barragán, & Boehm-Davis, 2016; Havik, Roberts, Van Hout, Schreuder, & Haverkort, 2009; Hopp, 2014; Just et al., 1996; Kim & Christianson, 2017; Medina, Callender, Brantmeier, & Schultz, 2017; Sagarra, 2017).

The results are also compatible with the findings of studies of following instructions that have reported the involvement of WM resources in following written or oral instructions/directions in children and adults (Engle, Carullo, & Collins, 1991; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008; Jaroslawska, Gathercole, Allen, & Holmes, 2016; Jaroslawska, Gathercole, & Holmes, 2018; Yang, Gathercole, & Allen, 2014; Yang, Allen, & Gathercole, 2016). The results are further consistent with a memory capacity account (Just & Carpenter, 1992) that predicts that there is a relationship between sentence processing and working memory capacity. Hence, some complex structures would be more difficult to be processed than others because more critical information is required to be held in working memory. This would impose an additional load on working memory resources, hindering retrieval and performance in individuals with poor working memory resources. However, it has been argued that working memory is not needed for syntactic parsing itself but only later post-interpretative processes (Caplan & Waters, 2013). Thus, parsing the temporally reversed sentences might not be difficult, but reversing the order of the mental event representations in one's mind may require extra resources. This would be equal in both languages. Such an explanation would fit with the finding that despite the main effects of both language and temporal order condition, there was no significant interaction between these variables in the bilinguals' data.

The significant main effect of working memory capacity was limited to the monolingual English group, but we also found a marginal correlation between working memory capacity and recall accuracy in the bilingual group. We speculate that the failure to detect a significant effect in the L2 speakers may have been due to range compression that could have made it less likely to find a strong correlation between WM span and recall accuracy in the L2 group. The L2 group in the current study were mostly graduate engineering and health sciences students with probably higher GPAs. The monolingual group had a lower group mean score, which means they had more of a range of scores. Thus, there was a better chance to find a correlation between WM scores and recall accuracy in following complex two-clause instructions. It is also possible that the bilinguals' WM resources were so good that the differences among them did not matter as most of them had enough resources.

Our results showed a correlation between English-based phonological short-term memory and ability to follow complex instructions in the combined group or monolinguals and L2 learners. Participants with better phonological memory abilities were able to remember and retrieve more critical information, as exhibited in higher recall accuracy in performing complex instructions. The role of phonological memory in following L2 complex instructions in bilinguals was consistent with the findings of studies exploring the role of phonological short-term memory in L2 acquisition and processing (Ellis, 1996; Kormos & Sáfár, 2008; Martin & Ellis, 2012; Service, 1992, 2012).

The correlation between phonological memory scores and ability to follow complex instructions in L2 in the bilingual data suggests that phonological memory was involved in remembering verbal instructions. Our results are consistent with the finding of previous studies that phonological short-term memory plays an important role in L2 vocabulary, grammar, and fluency (Ellis & Sinclair, 1996; French & O'Brien, 2008; Service, 1992), and that, regardless of the context, individuals with a better phonological memory ability make greater gains in L2 oral fluency development than those with poorer skills (O'Brien et al., 2007). O'Brien et al. (2007) argue that the relationship between phonological memory ability and language learning is not constrained by age but extends into adulthood, and also plays a role in second language acquisition. Thus, our study suggests that phonological memory may sometimes support task performance in addition to learning as has been reported before.

Some variables in the bilinguals' linguistic background influenced recall accuracy and were correlated with the ability to accurately follow linguistically complex instructions. Bilinguals who began acquiring their non-native language earlier in life exhibited superior comprehension and performance compared to those who acquired L2 later in life. The onset of L2 age of acquisition of both written and spoken language receptive and productive skills was significantly correlated with the recall accuracy in carrying out complex instructions. Our findings on the effect of L2 age of acquisition on remembering instructions with complex temporal clauses are consistent with the findings of studies that investigated the relationship between L2 AoA and morpho-syntactic processing (Sakai et al., 2009), phonological processing (Archila-Suerte et al., 2015), and cerebral activation during language production (Bloch et al., 2009). Thus, early acquisition of an additional language correlates with more efficient processing and performance in bilinguals.

However, the degree of L2 proficiency was another good predictor of bilinguals' performance. Our results demonstrated that highly skilled participants were more accurate in understanding and acting out instructions with semantically complex clauses. Many experiments suggest that proficiency plays a role in L2 processing and performance. Proficient L2 learners are more efficient in processing, have more accurate representations and responses and are able to demonstrate near-/native-like language performance (Bel et al., 2016; Keating, 2017; Rossi et al., 2017; Tanner et al., 2014). Thus, an increase in proficiency can attenuate the burden imposed by an L2, allowing near-/native-like processing and performance. Although research shows that the degree of proficiency interacts with variables such as WM resources and influences processing and task performance in L2/L3 (Coughlin & Tremblay, 2013; Hummel, 2009), L2 proficiency did not significantly modulate L2 WM span in the present study. This result is similar to the one reported by Service et al. (2002), i.e. that highly proficient L2 speakers' L2 WM is at the level of native speakers. Another reason could be that we used arithmetic operations instead of reading sentences as the processing task. This task is less dependent on language.

Bilinguals who arrived later in life in an English-speaking country and had been using the L2 less frequently had a lower recall accuracy than those who came to Canada at a younger age and had had more exposure to L2 input. While daily L1 use was negatively correlated with bilinguals' performance, daily L2 use was positively correlated with following complex instructions. Bilinguals who were using L2 more frequently in daily and academic activities were able to maintain and retrieve more critical information in comprehending and acting out complex sentences. Extensive experience with a target language and its frequent use is likely to influence L2 proficiency, resulting in more efficiency in language processing and performance. This explanation is consistent with the findings of the behavioral and electrophysiological studies that extended exposure to L2 input can alter processing mechanisms, reduce L1 transfer effects and ultimately result in a native- or near native-like processing in L2 learners (Frenck-Mestre, 2002; Kroll et al., 2015; Pliatsikas & Marinis, 2013).

Consistent with the finding of Clark and Clark (1968), in our study, the type of connective (before vs. after) used did not have a consistent or statistically significant effect, and did not interact with the much stronger and consistent effects of language and temporal order condition. Regardless of the type connective, participants always had superior recall accuracy and performance when the order of mention of the events was in line with the conceptual order. Unlike the small children previously studied, our participants were adults who had a robust linguistic knowledge of connectives and were able to correctly interpret temporal relations (Evers-Vermeul et al., 2017).

4.5 Conclusions

This empirical study reveals that the language of task (dominant vs. non-dominant), even in skilled L2 learners, affects bilingual processing and performance ability to recall sequences of complex instructions describing event order. There was a disadvantage for the tasks done in L2. A foreign language can put a strain on the computational and cognitive systems, deplete working memory resources and cause language processing and performance to have to rely more on available working memory capacity. In addition, when assessing WM capacity, the language of testing can modulate bilinguals' working memory span. Although we employed the operation span task with arithmetic operations rather than sentence comprehension, the language of testing influenced bilinguals' WM spans, with lower WM span scores exhibited in the non-dominant language.

We found that variables in bilinguals' language history predicted their performance in L2 tasks. The onset AoA of L2, functional proficiency, age of arrival in an L2 environment, language use, and exposure affected instruction recall accuracy as shown in the ability to remember and follow linguistically complex instructions. Thus, acquiring an additional language earlier in life, having better L2 skills, extensive exposure to L2 input, and frequent use of the second language resulted in efficient processing and superior execution in L2 learners. However, since language learning, development, and processing are dynamic processes (de Bot, 2012; De Bot et al., 2007), more experience with the target language can alter the processing mechanism and ultimately, result in efficiency in L2 performance. For example, most of these factors were correlated, L2 AoA and age of arrival: $r =.46$, L2 AoA and L2 proficiency: $r = -.40$, L2 proficiency and daily L2 use: $r = .43$, and L2 proficiency and exposure to L2 input: $r = .40$. This suggests that a better L2 performance reflects proficiency, and also automaticity with L2.

Our study further showed comprehension and performance difficulties related to the linguistic complexity of temporal-order sentences containing *before* and *after* adverbial connectives. Robust difficulties were observed in the execution of complex instructions when the order of mention of two events did not match their conceptual order. This suggested that the reverse condition imposed extra demands on the computational and cognitive systems. The influence of the linguistic load is further supported by the correlation between working memory capacity, phonological memory, and following the complex spoken instructions. Faced with demanding structures like the two-clause temporal order sentences, which required revising the order of clauses and modifying the mental representations, an advantage for participants with higher working memory abilities than those who have poorer working memory was revealed. This indicates the importance of available working memory resources for individuals to be able to simultaneously process the input and maintain the critical information for performing complex linguistic and cognitive tasks, like following the complex multi-step oral instructions in this study. The interactions between language or group with temporal order condition were not significant. Although this could mean that our tasks were not sensitive enough, the consistent main effects suggest a different possibility that the negative effect of temporal structure complexity was worse for the L2 than the L1 speakers.

The findings of this study have implications for the second language acquisition (SLA) theories, L1 and L2 processing and performance, and the use of a non-dominant language in the workplace, education, and daily activities. As our results show the disadvantage for recall accuracy in late bilinguals, the age of L2 acquisition should be included in data analysis. Managers, supervisors, and educators should be aware of the consequences of using a non-dominant language in the workplace and educational settings, where the English-only policy is at work. It might have real-world and cognitive consequences for non-native speakers, especially those with inferior second language skills. Non-native speakers might forget some of the critical information, fail to retain the sequential order of events, or maybe slower when they are supposed to follow instructions in their responsibilities.

5 Summary and Conclusion

5.1 Thesis Summary

The main goal of this thesis was to increase our understanding of the factors that influence encoding, decoding and executing real-world tasks, specifically the ability to follow spoken instructions in the second language of bilinguals. The critical questions in this thesis were (i) if the language of testing (L1 or L2) affects the ability to comprehend and perform sequences of spoken instructions, (ii) whether there is a relationship between individual differences in working memory capacity and phonological short-term memory with following sequential oral instructions, (iii) to what extent linguistic complexity modulates the comprehension and performance of multi-step verbal commands in L2, and (iv) which variables in bilinguals' language background predict L2 performance in understanding and executing a set of oral instructions.

To answer these questions, we designed and conducted three experiments consisting of tasks that mimicked real-life activities. In these, participants listened to sequences of oral instructions, and then either acted out the instructions or verbally repeated them back. This simulated a common phenomenon in real-life, where a manager or supervisor may require their employees to perform a sequence of duties, a teacher or professor may teach their students how to solve a problem or do an experiment by following certain steps or procedures, or a physician might instruct their patients how often and when to take specific medications.

Presenting target sentences in bilingual participants' first and second language allowed us to observe any difficulties in understanding and remembering oral instructions that were associated with the language of task (dominant or non-dominant) and the degree of proficiency in a non-native language. Furthermore, we presented the experimental materials orally as decoding natural-speed spoken language can be expected to be challenging for L2 learners and to overload working memory resources.

To test the hypothesis that WM measures correlate with the accuracy of recalling instructions, we investigated the relationship between working memory capacity and recall accuracy in following oral instructions by fitting the data into mixed effects models. We also explored if the language of presentation affects WM span in bilinguals' L1 and L2. To this aim, we presented a separate complex span task (i.e., the operation span task) in both languages of bilinguals. We further measured participants' phonological memory using a non-word repetition task with English-sounding nonwords. The relationship between phonological memory and bilinguals' memory for instructions in L2 was explored.

To investigate the relationship between the variables in bilinguals' language history, we collected information on participants' L2 AoA, the degree of functional proficiency, L1 and L2 use in different situations, and other factors. We fitted the data into mixed effects models to observe which factors predict bilinguals' performance in following spoken instructions in a non-dominant language.

Lastly, we manipulated the linguistic complexity of temporal order expressions in a set of instructions to test the hypothesis that some constructions, such as those that reverse the conceptual order of sequential events, are more difficult to process and perform. Our hypothesis was that if utterance complexity and L2 use loaded WM at the same time, complexity effects would be magnified for L2 stimuli. We fitted the data for remembering complex two-clause instructions into mixed effects models to observe the effect of linguistic complexity on recall accuracy and its interactions with the language of presentation and other independent variables.

The findings of the thesis are summarized below.

In all three experiments, regardless of the type of the task, enactment or verbal recall in Experiments 1 & 2 and enactment in Experiment 3, or task complexity, there was consistently a disadvantage for receiving the instructions in the L2. The recall accuracy in following spoken instructions was significantly lower when bilinguals were acting out or orally repeating sequences of instructions in the non-dominant language. Interestingly, the main effect of the language of testing was observed also when we measured participants' functional memory using a complex span task (i.e. the operation span task) in L1 and L2. Persian and Chinese bilingual participants exhibited lower WM span scores when the tasks were presented in the L2 (English). However, L1 and L2 WM spans were highly correlated. The inferior performance in an L2, even in the highly skilled Persian L2 learners, indicates that language comprehension and performance in a non-dominant language is demanding. It is associated with an additional load and might deplete working memory resources, causing the L2 task to be cognitively costly. The results are consistent with the findings of studies that processing input in a non-native language puts a strain on computational and cognitive systems and consumes working memory resources (Roberts, 2012; Service et al., 2002; van den Noort et al., 2006).

Our second finding was that individual differences in working memory capacity were correlated with the instruction-following ability. English monolinguals and those bilinguals who had greater working memory capacity demonstrated superior comprehension and performance. One reason for the main effect of WMC on following multi-step oral instructions can be that we used the complex span task to measure participants' WM span. The nature of most complex span tasks, such as the operation span task in our studies, is based on simultaneous processing (e.g., doing arithmetic operations) and storage (e.g., memorizing the words for memory after each arithmetic operation). The same processes are involved in following multi-step instructions. People have to encode the input, maintain the critical information in memory for later retrieval, remember the sequential order of each individual instruction in a given sequence, and then retrieve the information for execution or verbal recall. Furthermore, when the linguistic task was made harder in processing the reversed two-clause temporal order constructions in Experiment 3, they need to re-arrange the mental order of the events in the instructions to build a coherent mental representation. In the verbal recall task, they had to successfully verbally encode the action descriptions to reproduce the instructions after encoding, maintaining and retrieving the critical information.

Although we measured WMC separately in L1 and L2, we used only the L1 WM span to explore the main effect of WMC and its interactions with the other independent variables. The main reason was that although L1 and L2 WM spans were correlated (r = .22 in Experiment 1, r = .66 in Experiment 2, and r = .59 in Experiment 3), WM tested in L1 was assumed to provide a more valid estimate of individual WMC uncontaminated by L2 load. Our findings on the role of working memory resources in following instructions were consistent with previous studies that found a role for WMC in following instructions/directions in both children and adults (Engle et al., 1991; Gathercole et al., 2008; Jaroslawska et al., 2016, 2018; Yang et al., 2016; Yang et al., 2014). Thus, individuals with greater internal WM resources were less likely to forget the critical information or the sequential order in following multi-step oral instructions.

Our experimental method measuring participants' complex span in both L1 and L2 contributes to the ongoing debate on the relationship between memory span and the language of testing. Although bilinguals consistently exhibited greater WMC in their dominant language, L1 and L2 WMC were correlated. Thus, the notion of totally separate working memory resources in L2 was ruled out. Yet, the language of task (dominant or non-dominant language) resulted in a greater WM span in L1. Despite significant main effects of WM span on recall accuracy in

Experiments 1 and 2, the main effect of WM span on recall accuracy in following complex instructions was marginal in the bilingual group in Experiment 3. We manipulated linguistic complexity in Experiment 3, and it has been argued that WM is not needed for syntactic parsing itself but only for later post-interpretative processes (Caplan & Waters, 2013). Thus, syntactic parsing of the reversed sentences would be difficult but would not tax general WM. In contrast, reversing the mental event representation times in one's mind may have required extra WM resources, which would have been equal in both languages. This would fit with the finding that despite the main effects of both language and temporal order condition, there was no significant interaction between these variables in the bilinguals' data. We also speculate that range compression in the bilingual data and/or lower mean scores of monolinguals are other possible reasons for a non-significant correlation between WM scores and recall accuracy in following complex instructions in bilinguals in Experiment 3.

In Experiment 1, phonological short-term memory did not significantly affect the ability to follow sequences of simple instructions in skilled bilinguals. However, it was correlated with recall accuracy in following oral instructions when bilingual participants were less proficient (in Experiment 2) and also for advanced bilinguals when the task was linguistically complex (in Experiment 3). Thus, in

Experiments 2 and 3, individual differences in phonological memory were correlated with instruction-following ability. This indicates that the role of phonological memory in language learning and bilingual performance continues into adulthood; however, its impact can be modulated by the degree of L2 proficiency and task complexity.

The third set of findings relates to the circumstances of L2 learning that predict memory for instructions. Some of the background variables that we investigated, such as the onset of L2 age of acquisition and the degree of proficiency in English as a second language, were significant predictors of instruction-following ability in the L2. The effects of the AoA and proficiency on following verbal instructions were observed in less-skilled Chinese L1 bilinguals (Experiment 2) or for linguistically complex instructions in advanced Persian L1 bilinguals in Experiment 3. Faced with information overload and linguistic complexity, bilinguals who had had early exposure to the L2 of the task and/or had a higher language knowledge or skills showed superior comprehension and performance compared to those who had begun acquiring the L2 later in life or had reached a lower level of proficiency. Besides the L2 AoA and proficiency, other elements in bilinguals' language history, such as language use, exposure, and age of arrival in an L2 environment, also influenced information processing and performance for participants processing linguistically complex instructions in Experiment 3. It seems that processing more complex structures, e.g., temporal order constructions, or performing more complex tasks relies on more language skills, even in highly skilled bilinguals.

The fourth set of findings relates to coping with task difficulty. In Experiment 3, task complexity affected the ability to follow sequences of verbal instructions. Participants exhibited a lower recall accuracy in understanding and performing semantically complex instructions describing sequential temporal order. The recall accuracy reflected in instruction performance was affected by the congruency or incongruency between the order of mention of events and the conceptual order of events. Recall accuracy was significantly lower when the conceptual order did not match the linear order of mention. Thus, in the reverse condition, where the factual order was not consistent with the order of presentation, revising the order of events to construct a coherent mental representation before executing sequences of instructions overloaded available memory resources, resulting in a fragile representation and weaker retrieval. The inference of pronounced difficulties in re-arranging the sequential order of actions in temporal-order constructions was supported by a correlation between instruction performance and L1 WMC. Individuals with larger WM spans were probably more efficient at message processing and better at retrieval, which resulted in superior performance. This supports the domain-general unitary working memory approach. Thus, in the case of linguistic/task complexity, more working memory resources are needed, causing the processing and/or performance to be more challenging for individuals who do not have sufficient working memory.

Finally, in Experiments 1 and 2, the type of the task (performance vs. verbal recall) affected recall accuracy. Both monolingual and bilingual participants were consistently more accurate when they were acting out the sequences of spoken instructions rather than verbally recalling and repeating them. It seems that the verbal recall of action information relies more on working memory resources. Another reason could be that, in addition to encoding, the verbal recall task includes an output process in which participants produce the spoken instructions rather than execute them.

5.2 Conclusions

In conclusion, following spoken instructions, which involve information processing and performance, can be affected by a number of independent variables such as the language of task (dominant or non-dominant), individual differences in WMC, phonological memory, variation in language background, and the linguistic complexity in bilinguals. The non-dominant language or less-proficient language is associated with more pronounced difficulties in processing, comprehension, and performance. Variation in working and phonological memory abilities play an important role in memory for instructions. Thus, participants with less available working memory resources, or those whose working memory capacity is under a certain threshold, are more likely to forget the critical information or the sequential order of information. This suggests that sufficient working memory capacity is needed to carry out linguistic or information processing to successfully execute daily activities or professional responsibilities. The onset of L2 age of acquisition, the degree of L2 proficiency, and language use and exposure were shown to be among predictors of bilinguals' information processing and performance in following oral commands in a non-dominant language. The effect of the variables in bilinguals' L2 language background was more pronounced in less-skilled bilinguals or when the tasks were linguistically complex. The linguistic complexity of temporal order in two-clause constructions affects the comprehension and performance of multi-step instructions. It seems that revising the already constructed mental representation, when the conceptual order is incongruent with the order of presentation, is associated with robust difficulty and puts an extra load on the cognitive system, as exhibited by lower recall accuracy in the reverse condition in Experiment 3 of this study.

Although bilinguals demonstrated better understanding and performance when the tasks were presented in their dominant language, skilled L2 learners' performance was at the level of the English monolingual group. We did not find significant differences between the recall accuracy in following sequences of oral instructions in the proficient bilingual group and the monolingual control group. This shows that L2 processing and performance is a dynamic process, whose development can be extended into adulthood, and that late adult L2 learners may be able to demonstrate near-native-like linguistic and information processing and performance.

5.3 Limitations

One of the limitations of this study results from the sample of Persian L1 bilinguals. As most of our Persian speaking bilinguals were international graduate students, they had a higher average age than our control participants. The age factor, as reflected in the age of arrival in an L2 environment, affected the L2 performance in Experiment 3. However, further investigations by fitting the data into mixed effects regression models showed that the effects of other variables such as the initial L2 AoA, proficiency, exposure, and language use were more robust. Future studies should consider any modulating effects of the age of participants in tasks that require the active involvement of working memory ability and language background.

The other limitation was the unequal number of male and female participants in the monolingual group. As our native English speaker participants were recruited through the online linguistics student participant pool, this was inevitable.

5.4 Suggestions for future research

The ability to follow spoken instructions is a new area of research in bilingual studies. To our best knowledge, this is one of the first studies on memory for sequences of oral instructions. One follow-up possibility would be even more life-like simulation tasks that require participants to follow specific directions or instructions while doing a concurrent information processing task, a grammaticality judgment task or a semantic plausibility task. Participants could be involved in a GPS-like driving test and their performance and behavioral and/or electrophysiological responses would be collected. Our L2 participants reported that they were partially able to rehearse the information in the complex span task and the main instruction-following task in L1 rather than in L2. Designing experiments that include a secondary task that selectively disrupts the central executive as well as suppresses the phonological loop may reveal further difficulties associated with the extra loads a non-dominant language puts on the cognitive system. Another idea would be recruiting simultaneous or second/third generation bilingual participants to investigate the effect of L1 attrition on following spoken instructions in L2/L3 environment.

References

A collection of Chinese corpora and frequency lists. (n.d.). Retrieved from http://corpus.leeds.ac.uk/query-zh.html

Allen, R. J., & Waterman, A. H. (2015). How does enactment affect the ability to follow instructions in working memory? *Memory and Cognition*, *43*(3), 555–561. https://doi.org/10.3758/s13421-014-0481-3

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Memory and Language*, *11*(4), 417.

Ansell, B. J., & Flowers, C. R. (1982). Aphasic adults' understanding of complex adverbial sentences. *Brain and Language*, *15*(1), 82–91.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language and Hearing Research*, *49*, 970–983. https://doi.org/10.1044/1092-4388(2006/070)

Archila-Suerte, P., Zevin, J., & Hernandez, A. E. (2015). The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain and Language*, *141*, 35–49. https://doi.org/10.1016/j.bandl.2014.11.005

Assi, S. M. (n.d.). Persian Linguistic Database (PLDB). Retrieved from http://pldb.ihcs.ac.ir/

Baayen, R. H. (2008). *Analyzing Linguistic Data*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. (2007). *Working Memory, Thought, and Action*. *Oxford Psychology Series* (Vol. 45). Oxford: OUP.

https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies - annurev-psych-120710-100422. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D. (2017). Working memory: Theories, models, and controversies. *Exploring Working Memory: Selected Works of Alan Baddeley*, 333–369. https://doi.org/10.4324/9781315111261

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv Preprint ArXiv:1406.5823*.

Bel, A., Sagarra, N., Comínguez, J. P., & García-Alcaraz, E. (2016). Transfer and proficiency effects in L2 processing of subject anaphora. *Lingua*, *184*, 134–159. https://doi.org/10.1016/j.lingua.2016.07.001

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009

Bloch, C., Kaiser, A., Kuenzli, E., Zappatore, D., Haller, S., Franceschini, R., … Nitsch, C. (2009). The age of second language acquisition determines the variability in activation elicited by narration in three languages in Broca's and Wernicke's area. *Neuropsychologia*, *47*(3), 625–633. https://doi.org/10.1016/j.neuropsychologia.2008.11.009

Blything, L. P., & Cain, K. (2016). Children's Processing and Comprehension of Complex Sentences Containing Temporal Connectives : The Influence of Memory on the Time Course of Accurate Responses, *52*(10), 1517–1529. https://doi.org/http://dx.doi.org/10.1037/dev0000201

Blything, L. P., Davies, R., & Cain, K. (2015). Young Children ' s Comprehension of Temporal Relations in Complex Sentences : The Influence of Memory on Performance, *86*(6), 1–17. https://doi.org/10.1111/cdev.12412

Brandt, S., Nitschke, S., & Kidd, E. (2017). Priming the Comprehension of German Object Relative Clauses. *Language Learning and Development*, *13*(3), 241–261. https://doi.org/10.1080/15475441.2016.1235500

Briellmann, R. S., Saling, M. M., Connell, A. B., Waites, A. B., Abbott, D. F., & Jackson, G. D. (2004). A high-field functional MRI study of quadri-lingual subjects. *Brain and Language*, *89*(3), 531–542. https://doi.org/10.1016/j.bandl.2004.01.008

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Caplan, D., & Waters, G. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin and Review*, *20*(2), 243–268. https://doi.org/10.3758/s13423-012-0369-9

Charlesworth, L. A., Allen, R. J., Morson, S., Burn, W. K., Souchay, C., Charlesworth, L. A., … Souchay, C. (2014). Working Memory and the Enactment Effect in Early Alzheimer's Disease. *ISRN Neurology*, *2014*, 1–5. https://doi.org/10.1155/2014/694761

Christiansen, M. H., & Chater, N. (2016). *Creating language. Integrating evolution, acquisition, and processing*. *MIT Press*. https://doi.org/10.1017/CBO9781107415324.004

Clark, H. H., & Clark, E. V. (1968). Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology*, *20*(2), 129–138. https://doi.org/10.1080/14640746808400141

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266–275.

Conway, A. R. a., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychon Bull Rev*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*(3), 615–646. https://doi.org/10.1017/S0142716411000890

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2005). Working-Memory Capacity Limits in a Theoretical Context. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application: The 4th Tsukuba International Conference on Memory* (p. 155–175). Lawrence Erlbaum Associates Publishers.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and Peripheral Components of Working Memory Storage. *Journal of Experimental Psychology*, *143*(5), 1806–1836. https://doi.org/10.1037/a0036814

Dai, Y. (2015). Working Memory in L2 Sentence Processing: The Case with Relative Clause Attachment. In A. M. Z. Wen, M. B. Mota (Ed.), *Working Memory in Second Language Acquisition and Processing* (pp. 105–124). NY: Multilingual Matters.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory during reading. *Journal Of Verbal Learning And Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

De Bleser, R., Dupont, P., Postler, J., Bormans, G., Speelman, D., Mortelmans, L., & Debrock, M. (2003). The organisation of the bilingual lexicon: A PET study. *Journal of Neurolinguistics*, *16*(4–5), 439–456. https://doi.org/10.1016/S0911-6044(03)00022-8

de Bot, K. (2012). Rethinking multilingual processing. *Third Language Acquisition in Adulthood*, *46*, 79. https://doi.org/10.1075/sibil.46.07bot

De Bot, K., & Jaensch, C. (2015). What is special about L3 processing? *Bilingualism*, *18*(2), 130–144. https://doi.org/10.1017/S1366728913000448

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

DeKeyser, R. M. (2013). Age Effects in Second Language Learning: Stepping

Stones Toward Better Understanding. *Language Learning*, *63*(SUPPL. 1), 52–67. https://doi.org/10.1111/j.1467-9922.2012.00737.x

DeKeyser, R. M. (2017). Age in Learning and Teaching Grammar. In T. J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). International Association and M. DelliCarpini. https://doi.org/10.1002/9781118784235.eelt0106

Delcenserie, A., & Genesee, F. (2017). The effects of age of acquisition on verbal memory in bilinguals. *International Journal of Bilingualism*, *21*(6), 600–616. https://doi.org/10.1177/1367006916639158

Deng, T., Dunlap, S., & Chen, B. (2017). Effects of input training on second language syntactic representation entrenchment. *International Journal of Bilingualism*, *21*(1), 3–20. https://doi.org/10.1177/1367006915589423

Diessel, H. (2004). *The acquisition of complex sentences* (Vol. 105). Cambridge University Press.

Dornic, S. (1980). Information processing and language dominance 1. *Applied Psychology*, *29*(1-2), 119–140. https://doi.org/10.1111/j.1464-0597.1980.tb00886.x

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, *26*(4), 443–472. https://doi.org/10.1177/0267658310373326

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish - English bilinguals. *Bilingualism*, *10*(1), 101–116. https://doi.org/10.1017/S1366728906002847

Ellis, N. C. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 234–250.

Engle, R., Carullo, J., & Collins, K. (1991). Individual differences in working memory for comprehension and following directions Cités par moi. *The Journal of Educational Research*, *84*(5), 253–262.

Evers-Vermeul, J., Hoek, J., & Scholman, M. C. J. (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue and*

Discourse, *8*(2), 1–20. https://doi.org/10.5087/dad.2017.201

Farmer, T. A., Fine, A. B., Misyak, J. B., & Christiansen, M. H. (2017). Reading span task performance, linguistic experience, and the processing of unexpected syntactic events. *Quarterly Journal of Experimental Psychology*, *70*(3), 413–433. https://doi.org/10.1080/17470218.2015.1131310

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish* (Vol. 29). John Benjamins Publishing.

Foroughi, C. K., Barragán, D., & Boehm-Davis, D. A. (2016). Interrupted reading and working memory capacity. *Journal of Applied Research in Memory and Cognition*, *5*(4), 395–400.

Francis, A. L., Tigchelaar, L. J., Zhang, R., & Zekveld, A. A. (2018). Effects of Second Language Proficiency and Linguistic Uncertainty on Recognition of Speech in Native and Nonnative Competing Speech. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1815–1830.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*(3), 463–487. https://doi.org/10.1017/S0142716408080211

Frenck-Mestre, C. (2002). 9 An on-line look at sentence processing in the second language. *Advances in Psychology*, *134*(C), 217–236. https://doi.org/10.1016/S0166-4115(02)80012-7

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, *22*(8), 1019–1037. https://doi.org/10.1002/acp.1407

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working Memory in the Classroom. In S. J. Pickering (Ed.), *Educational Psychology, Working Memory and Education* (pp. 219–240). Oxford: Elsevier Academic Press. https://doi.org/10.1016/B978-012554465-8/50010-7

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological Memory and Vocabulary Development During the Early School Years: A Longitudinal Study. *Developmental Psychology*, *28*(5), 887. https://doi.org/10.1037/0012-1649.28.5.887

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67. https://doi.org/10.1017/S1366728998000133

Harrington, M., & Sawyer, M. (1992). L2 Working Memory Capacity and L2 Reading Skill. *Studies in Second Language Acquisition*, *14*(1), 25–38. https://doi.org/10.1017/S0272263100010457

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112.

Ho, Y. (2002). *Chinese-English frequency dictionary: a study guide to Mandarin Chinese's 500 most frequently used words*. Hippocrene Books.

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition*, *21*(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnative adult learners. *Applied Psycholinguistics*, *30*, 225–249. https://doi.org/10.1017/S0142716409090109

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/J.JML.2007.11.007

Jaroslawska, A. J., Gathercole, S. E., Allen, R. J., & Holmes, J. (2016). Following instructions from working memory: Why does action at encoding and recall help? *Memory and Cognition*, *44*(8), 1183–1191. https://doi.org/10.3758/s13421-016-0636-5

Jaroslawska, A. J., Gathercole, S. E., & Holmes, J. (2018). Following instructions in a dual-task paradigm: Evidence for a temporary motor store in working memory. *Quarterly Journal of Experimental Psychology*, 174702181774349. https://doi.org/10.1177/1747021817743492

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension:

individual differences in working memory. *Psychological Review*, *99*(1), 122. https://doi.org/10.1037/0033-295X.99.1.122

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review*, *103*(4), 773–780. https://doi.org/10.1037/0033-295X.103.4.773

Keating, G. D. (2017). L2 Proficiency matters in comparative L1/L2 processing research. *Bilingualism*, *20*(4), 700–701. https://doi.org/10.1017/S1366728916000912

Kilman, L., Zekveld, A., Hällgren, M., & Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology*, *5*, 651.

Kim, J. H., & Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Language Research*, *33*(3), 365–388. https://doi.org/10.1177/0267658315623322

Koriat, A., Ben-Zur, H., & Nussbaum, A. (1990). Encoding information for future action: Memory for to-be-performed tasks versus memory for to-be-recalled tasks. *Memory & Cognition*, *18*(6), 568–578. https://doi.org/10.3758/BF03197099

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

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

Liang, L., & Chen, B. (2014). Processing morphologically complex words in second-language learners: The effect of proficiency. *Acta Psychologica*, *150*, 69–79. https://doi.org/10.1016/j.actpsy.2014.04.009

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lui, S. S. Y., Yang, T. X., Ng, C. L. Y., Wong, P. T. Y., Wong, J. O. Y., Ettinger, U., … Chan, R. C. K. (2018). Following Instructions in Patients with Schizophrenia: The Benefits of Actions at Encoding and Recall. *Schizophrenia Bulletin*, *44*(1), 137–146. https://doi.org/10.1093/schbul/sbx026

Macnamara, B. N., & Conway, A. R. A. (2016). Working Memory Capacity as a Predictor of Simultaneous Language Interpreting Performance. *Journal of Applied Research in Memory and Cognition*, *5*(4), 434–444. https://doi.org/10.1016/j.jarmac.2015.12.001

Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity*, *21*, 41–49. https://doi.org/10.1016/j.tsc.2016.05.005

Mandler, J. M. (1986). On the comprehension of temporal order. *Language and Cognitive Processes*, *1*(4), 309–320. https://doi.org/10.1080/01690968608404680

Marsh, L. G., & Maki, R. H. (1976). Efficiency of arithmetic operations in bilinguals as a function of language. *Memory & Cognition*, *4*(4), 459–464. https://doi.org/10.3758/BF03213203

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(3), 379–413. https://doi.org/10.1017/S0272263112000125

McClain, L., & Huang, J. Y. S. (1982). Speed of simple arithmetic in bilinguals. *Memory & Cognition*, *10*(6), 591–596. https://doi.org/10.3758/BF03202441

Medina, A., Callender, A. A., Brantmeier, C., & Schultz, L. (2017). Inserted adjuncts, working memory capacity, and L2 reading. *System*, *66*, 69–86. https://doi.org/10.1016/j.system.2017.03.002

Meschyan, G., & Hernandez, A. E. (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: An fMRI investigation. *NeuroImage*, *29*(4), 1135–1140. https://doi.org/10.1016/j.neuroimage.2005.08.055

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, *130*(2), 163–

168. https://doi.org/10.1037//0096-3445.130.2.163

Miyake, A., & Priti, S. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174909

Morgan-Short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PLOS ONE*, *7*(3), e32974. https://doi.org/10.1371/journal.pone.0032974

Natsopoulos, D., & Abadzi, H. (1986). Understanding Linguistic Time Sequence and Simultaneity : A Literature Review and Some New Data. *Journal of Psycholinguistic Research. 15*(3), 243–244.

Natsopoulos, D., Mentenopoulos, G., Bostantzopoulou, S., Katsarou, Z., Grouios, G., & Logothetis, J. (1991). Understanding of relational time terms before and after in Parkinsonian patients. *Brain and Language*, *40*(4), 444–458.

Neeley, T. B. (2013). Language Matters: Status Loss and Achieved Status Distinctions in Global Organizations. *Organization Science*, *24*(2), 319–644. https://doi.org/10.1287/orsc.1120.0739

O'brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, *29*(4), 557–581. https://doi.org/https://doi.org/10.1017/S027226310707043X

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *14*(7), 758–799. https://doi.org/10.1037/bul0000046

Oron, A., Szymaszek, A., & Szelag, E. (2015). Temporal information processing as a basis for auditory comprehension : clinical evidence from aphasic patients. *International journal of language & communication disorders*, 604–615. https://doi.org/10.1111/1460-6984.12160

Osaka, M., & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*(4), 287–289. https://doi.org/10.3758/BF03330466

Osaka, M., Osaka, N., & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society*, *31*(2), 117–118. https://doi.org/10.3758/BF03334156

Paap, K. R., Sawi, O. M., Dalibar, C., Darrow, J., & Johnson, H. A. (2015). Beyond Panglossian Optimism: Larger N2 Amplitudes Probably Signal a Bilingual Disadvantage in Conflict Monitoring. *AIMS Neuroscience*, *2*(1), 1–6. https://doi.org/10.3934/Neuroscience.2015.1.1

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *34*(4), 388.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., … Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48. https://doi.org/10.1037/bul0000124. Epub 2017 Oct 30

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206. https://doi.org/10.1016/J.CONB.2005.03.007

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., … Mehler, J. (1998). The bilingual brain : Proficienty and age of acquisition of the second language. *Brain*, *121*(10), 1841–1852. https://doi.org/10.1093/brain/121.10.1841

Pliatsikas, C., & Marinis, T. (2013). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics*, *34*(5), 943–970. https://doi.org/10.1017/S0142716412000082

Politzer-ahles, S., Xiang, M., & Almeida, D. (2017). " Before " and " after ": Investigating the relationship between temporal connectives and chronological ordering using event- related potentials, 1–17. https://doi.org/10.1371/journal.pone.0175199

Price, J., Catrambone, R., & Engle, R. W. (2007). When capacity matters: the role of working memory in problem solving. In D. H. Jonassen (Ed.), *Learning to Solve Complex Scientific Problems* (pp. 49–76). NY: Lawrence Erlbaum.

R Core Team. (2017). R: A language and environment for statistical com-puting.

R Foundation for Statistical Computing, Vienna, Austria. URL https://www. R-project. org.

Reichle, R. V, Tremblay, A., & Coughlin, C. E. (2013). Working-memory capacity effects in the processing of non-adjacent subject-verb agreement: An event-related brain potentials study. *Selected Proceedings of the 2011 Second Language Research Forum*, 54–69.

Roberts, L. (2012). Individual differences in second language sentence processing. *Studies in Second Language Acquisition*, *13*(02), 275. https://doi.org/10.1177/0267658311418416

Roncaglia-Denissen, M. P., & Kotz, S. A. (2016). What does neuroimaging tell us about morphosyntactic processing in the brain of second language learners? *Bilingualism*, *19*(4), 665–673. https://doi.org/10.1017/S1366728915000413

Rosselli, M., Ardila, A., Lalwani, L. N., & Vélez-Uribe, I. (2016). The effect of language proficiency on executive functions in balanced and unbalanced Spanish-English bilinguals. *Bilingualism*, *19*(3), 489–503. https://doi.org/10.1017/S1366728915000309

Rossi, E., Diaz, M., Kroll, J. F., & Dussias, P. E. (2017). Late bilinguals are sensitive to unique aspects of second language processing: Evidence from clitic pronouns word-order. *Frontiers in Psychology*, *8*, 342. https://doi.org/10.3389/fpsyg.2017.00342

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism*, *17*(3), 542–555. https://doi.org/10.1017/S1366728913000643

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, *33*(3), 341–363. https://doi.org/10.1177/0267658317690577

Sakai, K. L., Nauchi, A., Tatsuno, Y., Hirano, K., Muraishi, Y., Kimura, M., … Yusa, N. (2009). Distinct roles of left inferior frontal regions that explain individual differences in second language acquisition. *Human Brain Mapping*, *30*(8), 2440–2452. https://doi.org/10.1002/hbm.20681

Science and Technology Facilities Council. (2016). MRC Psycholinguistic Database. *Http://Www.Psych.Rl.Ac.Uk/, accessed D*, 500–502.

Service, E. (1992). Phonology, Working Memory, and Foreign-language Learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21–50. https://doi.org/10.1080/14640749208401314

Service, E. (2012). Working Memory in Second Language Acquisition: Phonological Short-Term. *The Encyclopedia of Applied Linguistics*, 1–5. https://doi.org/10.1002/9781405198431.wbeal1287

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*(3), 383–408. https://doi.org/10.1080/09541440143000140

Service, E., & Tujulin, A. M. (2002). Recall of morphologically complex forms is affected by memory task but not dyslexia. *Brain and Language*, *81*(1–3), 42–54. https://doi.org/10.1006/brln.2001.2505

Smyth, M. M., & Pendleton, L. R. (1989). Working Memory for Movements. *The Quarterly Journal of Experimental Psychology Section A*, *41*(2), 235–250. https://doi.org/10.1080/14640748908402363

Smyth, M. M., & Pendleton, L. R. (1990). Space and Movement in Working Memory. *The Quarterly Journal of Experimental Psychology Section A*, *42*(2), 291–304. https://doi.org/10.1080/14640749008401223

Sörqvist, P., Hurtig, A., Ljung, R., & Rönnberg, J. (2014). High second-language proficiency protects against the effects of reverberation on listening comprehension. *Scandinavian Journal of Psychology*, *55*(2), 91–96. https://doi.org/10.1111/sjop.12115

Takano, Y., & Noda, A. (1993). A Temporary Decline of Thinking Ability During Foreign Language Processing. *Journal of Cross-Cultural Psychology*, *24*(4), 445–462. https://doi.org/10.1177/0022022193244005

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism*, *17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

https://doi.org/10.1016/0749-596X(89)90040-5

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

van den Noort, M., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289–296. https://doi.org/10.1027/1016-9040.11.4.289

van den Noort, M., Struys, E., Kim, K., Bosch, P., Mondt, K., van Kralingen, R., … van de Craen, P. (2014). Multilingual processing in the brain. *International Journal of Multilingualism*, *11*(2), 182–201. https://doi.org/10.1080/14790718.2013.791298

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11–12), 943–953. https://doi.org/10.1016/j.specom.2010.05.002

van Hell, J. G. (2010). Event-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 proficiency levelsEvent-related brain potentials and second language learning: syntactic processing in late L2 learners at different L2 p. *Second Language Research*, *26*(1), 43.

Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., & Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia*, *101*, 17–29. https://doi.org/10.1016/j.neuropsychologia.2017.05.009

Van Rinsveld, A., Schiltz, C., Brunner, M., Landerl, K., & Ugen, S. (2016). Solving arithmetic problems in first and second language: Does the language context matter? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2016.01.003

Vejnović, D., Milin, P., & Zdravković, S. (2010). Effects of proficiency and age of language acquisition on working memory performance in bilinguals. *Psihologija*, *43*(3), 219–232. https://doi.org/10.2298/PSI1003219V

Volk, S., Köhler, T., & Pudelko, M. (2014). Brain drain: The cognitive neuroscience of foreign language processing in multinational corporations. *Journal of International Business Studies*, *45*(7), 862–885. https://doi.org/10.1057/jibs.2014.26

Wang, Y., Lin, L., Kuhl, P., & Hirsch, J. (2007). Mathematical and Linguistic Processing Differs Between Native and Second Languages: An fMRI Study. *Brain Imaging and Behavior*, *1*(3–4), 68–82. https://doi.org/10.1007/s11682-007-9007-y

Waterman, A. H., Atkinson, A. L., Aslam, S. S., Holmes, J., Jaroslawska, A., & Allen, R. J. (2017). Do actions speak louder than words? Examining children's ability to follow instructions. *Memory and Cognition*, *45*(6), 877–890. https://doi.org/10.3758/s13421-017-0702-7

Wattendorf, E., & Festman, J. (2008). Images of the multilingual brain: The effect of age of second language acquisition. *Annual Review of Applied Linguistics*, *28*, 3–24. https://doi.org/10.1017/S0267190508080033

Wattendorf, E., Festman, J., Westermann, B., Keil, U., Zappatore, D., Franceschini, R., … Nitsch, C. (2014). Early bilingualism influences early and subsequently later acquired languages in cortical regions representing control functions. *International Journal of Bilingualism*, *18*(1), 48–66. https://doi.org/10.1177/1367006912456590

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002

Yang, T., Allen, R. J., & Gathercole, S. E. (2016). Examining the role of working memory resources in following spoken instructions. *Journal of Cognitive Psychology*, *28*(2), 186–198. https://doi.org/10.1080/20445911.2015.1101118

Yang, T., Gathercole, S. E., & Allen, R. J. (2014). Benefit of enactment over oral repetition of verbal instruction does not require additional working memory during encoding. *Psychonomic Bulletin and Review*, *21*(1), 186–192.

https://doi.org/10.3758/s13423-013-0471-7

Yang, T. X., Allen, R. J., Holmes, J., & Chan, R. C. K. K. (2017). Impaired memory for instructions in children with attention-deficit hyperactivity disorder is improved by action at presentation and recall. *Frontiers in Psychology*, *8*(JAN), 39. https://doi.org/10.3389/fpsyg.2017.00039

Yang, T. xiao, Jia, L. xia, Zheng, Q., Allen, R. J., & Ye, Z. (2018a). Forward and backward recall of serial actions: Exploring the temporal dynamics of working memory for instruction. *Memory and Cognition*, 1–13. https://doi.org/10.3758/s13421-018-0865-x

Yang, T. xiao, Jia, L. xia, Zheng, Q., Allen, R. J., & Ye, Z. (2018b). Forward and backward recall of serial actions: Exploring the temporal dynamics of working memory for instruction. *Memory and Cognition*. https://doi.org/10.3758/s13421-018-0865-x

Ye, Z., Milenkova, M., Mohammadi, B., Kollewe, K., Schrader, C., Dengler, R., … Münte, T. F. (2012). Impaired comprehension of temporal connectives in Parkinson’s disease-A neuroimaging study. *Neuropsychologia*, *50*(8), 1794–1800. https://doi.org/10.1016/j.neuropsychologia.2012.04.004

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162.

Appendices

Appendix A

English Stimuli, List 1, Experiment 1 & 2

Sequences 1

a. place the red stapler in the box
b. put the highlighter into the blue tray
c & d. sign the tax form, and then put it into the yellow folder
e. put the black pen into the basket

Sequences 2

a & b. copy the sales report, and then put it into the red folder
c. put the highlighter into the yellow box
d. move the mouse onto the blue pad
e. place the red pen in the tray

Sequences 3

a. put the pen into the black tray
b. put the order form into the yellow folder
c. place the red marker in the box
d & e. pick up the green stapler, and then put it into the basket

Sequences 4

a. move the yellow mouse onto the pad
b & c. print the budget report, and then put it into the blue folder
d. place the ruler in the round basket
e. put the green highlighter into the tray

Sequences 5

a & b. pick up the green marker, and then put it into the blue tray
c. move the mouse onto the black pad
d. put the pen into the yellow box
e. place the stapler in the square basket

Sequences 6

a. place the marker in the red box
b. put the yellow highlighter into the basket
c & d. print the salary report, and then put it into the green folder
e. put the pen into the blue tray

Sequences 7

a. move the green mouse onto the pad
b & c. pick up the yellow marker, and then put it into the round basket
d. place the tax form in the blue folder
e. put the ruler into the red box

Sequences 8

a. put the salary report into the black folder
b. place the blue pen in the tray
c & d. pick up the highlighter, and then put it into the red box
e. put the marker into the square basket

Sequences 9

a. place the pen in the round basket
b. put the ruler into the red box
c. move the blue mouse into the tray
d & e. copy the application form, and then put it into the yellow folder

Sequences 10

a & b. sign the application form, and then put it into the black folder
c. move the red mouse into the basket
d. place the pen in the green box
e. put the blue marker into the tray

Sequences 11

a & b. pick up the blue stapler, and then put it into the box
c. put the order form into the red folder
d. place the marker in the black tray
e. move the mouse into the round basket

Sequences 12

a. put the red mouse into the box
b. place the budget report in the green folder
c & d. pick up the black ruler, and then put it into the tray
e. put the blue marker into the basket

Appendix B

Language background questionnaire

Language Background Questionnaire – Experiment 3

Participant information- *To be filled by the researcher*
Project name/code:___________ Participant code:_____________ Date:___________

Please answer the following questions to the best of your knowledge.

A. Demographic and Education Information

1. Age: [ ] 2. Gender: M◯ F◯

3. Education: Undergraduate ◯ Master's ◯ PhD ◯ Other ◯ 4. Program of study: ____________________

5. Have you ever had any perceptual (color blindness) problems? Yes◯ No◯

B. Language Acquisition & Proficiency

1. Your native language is: __________
2. Your second language is: _________ Additional languages (specify): ____________________________
3. Your Length of residence in an English speaking country: Canada, ____ Years ____ Months; other (specify) _______
4. Your age at which you first started to learn the second language & the ways you learned the 2nd language were:

Language ⬇	**Age first learned the language skills**				**Ways and years you learned the language**		**Total number of years you spent learning the language**
	Listening	**Speaking**	**Reading**	**Writing**	**Instruction** (Formal Teaching)	**Exposure** (Not Teaching)	
English							
Persian							
....							

5. Please list all languages you know, then rate your current language abilities according to the following scale. (**Circle the numbers in the table**).

1 ***very low*** **2** ***low*** **3** ***intermediate*** **4** ***advanced*** **5** ***near-native*** **6** ***native-like***

Language ⬇	**Reading proficiency**	**Writing proficiency**	**Speaking fluency**	**Listening ability**	**Overall Proficiency**
English	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6
Persian	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6
....	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6
....	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6	1 2 3 4 5 6

6. Please indicate the scores for any standardized test of language proficiency you have taken so far.

Language Test	Year	Listening	Speaking	Reading	Writing	Overall
IELTS						
TOEFL						
………….						

C. Language Use & Preference

1. What language(s) do you usually use to speak to your family members or roommates at home? ______________

2. Estimate, in terms of percentage, how often you use your native language and other languages per week (in all daily activities combined). *Total should equal to 100%.

Native language ________% English ______% other languages ________%

3. Which language do you use for the following activities? Please rate your <u>language use</u> according to the following scale.

1	**2**	**3**	**4**	**5**	**6**	**7**
Never	**Rarely**	**Occasionally**	**Sometimes**	**Frequently**	**Very Frequently**	**Always**

Language	**Doing arithmetics (add, multiply, count, and etc.)**	**Remembering or telling numbers (e.g., student ID, phone numbers)**	**Dreaming**	**Thinking**	**Talking to yourself**	**Expressing anger or affection**
English	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
Persian	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
………..	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
…………	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

4. Which language do you prefer to use for the following situations? Please rate your <u>language preference</u> according to the following scale.

1	**2**	**3**	**4**	**5**	**6**	**7**
Never	**Rarely**	**Occasionally**	**Sometimes**	**Frequently**	**Very Frequently**	**Always**

Language	**At home**	**At university**	**At a party**	**Non-Academic Reading**	**Watching TV & Listening to the Radio/Music**	**Social media**
English	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
Persian	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
………..	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7
…………	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7	1 2 3 4 5 6 7

5. Please add any additional relevant information to your language history.

Appendix C

Pseudo words used in the non-word repetition task

Non-word	Number of Syllables
ballop	2
bannow	2
diller	2
glistow	2
hampent	2
pennel	2
prindle	2
rubid	2
sladding	2
tafflest	2
bannifer	3
barrazon	3
brasterer	3
commerine	3
doppelate	3
frescovent	3
glistering	3
skiticult	3
thickery	3
trumpetine	3
blonterstaping	4
commeecitate	4
contramponist	4
empliforvent	4
fenneriser	4
loddenapish	4
penerriful	4
perplisteronk	4
stopograttic	4
woogalamic	4

Non-word	Number of Syllables
altupatory	5
confrantually	5
defermication	5
detratapillic	5
pristoractional	5
reutterpation	5
sepretennial	5
underbrantuand	5
versatrationist	5
voltularity	5

Appendix D

Stimuli of the English Operation Span Task

Sequence	Trial	Equation	WORD
1	1	(10/5) - 1 = 2	BEACH
	2	(2x5) + 2 = 12	PLATE
2	1	(8/2) + 3 = 7	CROWD
	2	(6x6) - 4 = 30	PANTS
	3	(9/3) - 1 = 4	CHEST
	4	(6x2) - 4 = 8	TRUCK
	5	(7x4) + 3 = 34	POWER
	6	(5/5) + 7 = 8	TEST
3	1	(2x6) + 4 = 16	WORLD
	2	(8/4) - 1 = 2	STEAK
	3	(5x8) - 4 = 36	GRASS
4	1	(7x6) - 2 = 40	PAINT
	2	(8/2) + 10 = 15	SWEAT
	3	(10/2) - 1 = 4	GLASS
	4	(7x7) - 5 = 41	STICK
	5	(9x2) + 2 = 20	FENCE
5	1	(10/2) - 2 = 3	DUST
	2	(8x4) - 5= 29	FLOOD
	3	(6/3) + 5 = 7	CLOCK
6	1	(7x8) - 5 = 51	CHILD
	2	(6/6) + 7 = 9	CHALK
	3	(3x9) + 2 = 28	BRAIN
	4	(9/3) + 2 = 5	CLOUD
7	1	(10/2) + 6 = 11	CHEEK
	1	(4x5) - 7 = 11	STEEL
	3	(8x2) + 6 = 22	COAST
	4	(6/2) + 7 = 9	HAND
8	1	(3x7) - 5 = 16	STORM
	2	(4/2) + 7 = 10	DRUG
	3	(8x7) - 6 = 50	BREAD
	4	(2x7) - 5 = 8	COACH

Sequence	Trial	Equation	WORD
8	5	(9/9) + 7 = 8	TRAIN
	6	(6/3) + 4 = 6	DESK
9	1	(7x4) + 4 = 31	BAND
	2	(6/2) - 1 = 2	SMOKE
	3	(4x6) - 8 = 16	QUEEN
10	1	(5x3) - 6 = 9	BRUSH
	2	(7/7) + 5 = 8	PLANT
	3	(4x4) + 5 = 20	CLUB
	4	(3x4) - 5 = 7	FILM
	5	(8/2) + 3 = 6	GIFT
11	1	(8/2) + 2 = 7	POST
	2	(6x8) - 6 = 42	SKIN
12	1	(4/4) + 5 = 7	CHAIN
	2	(3x8) - 4 = 20	WIND
	3	(6x6) + 5 = 40	PLANE
	4	(8/2) - 3 = 1	WAIST
13	1	(5x9) + 2 = 47	SMILE
	2	(8/2) + 5 = 7	CHART
14	1	(4/2) + 5 = 6	GOLD
	2	(5x5) - 2 = 23	CLOTH
	3	(8x2) + 9 = 24	DANCE
	4	(9/3) - 1 = 2	LUNCH
	5	(8/2) + 6 = 10	SNAKE
15	1	(4x4) - 5 = 11	FRUIT
	2	(8/8) + 6 = 9	DRESS
	3	(7x8) + 2 = 60	STONE
	4	(3x9) - 3 = 24	CROSS
	5	(10/2) - 4 = 2	LAND
	6	(2x8) + 3 = 19	BLOOD

Appendix E

A sample of a computer screen of a sequence of instructions in the enactment task in Experiment 1 & 2

Appendix F

English Stimuli, List 1, Experiment 3

Sequence 1. Drink the syrup, *then before you use the spray, take the tablet,* and last, administer eye drops.

Sequence 2. Stamp the transcript, *then after you sign the report, print the order form,* and last, copy the contract.

Sequence 3. Fax the licence, *then scan the complaint form, after you punch the schedule,* and last, file the manual.

Sequence 4. Take the capsule, *then administer nasal drops, before you use the spray,* and last, drink the syrup.

Sequence 5. Sign the proposal, *then after you punch the resume, stamp the paycheque,* and last, file the claim.

Sequence 6. Use the spray, *then before you drink the syrup, take the capsule,* and last, administer ear drops.

Sequence 7. Print the declaration, *then scan the pay stub, before you fax the lease,* and last, punch the checklist.

Sequence 8. Use the spray, *then drink the syrup, after you administer nasal drops,* and last, take the tablet.

Sequence 9. Fax the pay stub, *then before you stamp the budget form, scan the contract,* and last, sign the statement.

Sequence 10. Punch the declaration, *then after you copy the questionnaire, file the order form,* and last, print the record.

Sequence 11. Sign the paycheque, *then fax the memo, after you copy the licence,* and last, punch the notice.

Sequence 12. Administer eye drops, *then take the capsule, before you drink the syrup,* and last, use the spray.

Sequence 13. Print the notice, *then after you sign the schedule, copy the claim*, and last, stamp the declaration.

Sequence 14. Scan the record, *then before you fax the statement, print the complaint form,* and last, punch the lease.

Sequence 15. Sign the claim, *then copy the agreement*, *before you punch the report*, and last, fax the invoice.

Sequence 16. Drink the syrup, *then use the spray*, *after you take the tablet,* and last, administer ear drops.

Sequence 17. Copy the diploma, *then before you print the certificate, sign the invoice*, and last, stamp the record.

Sequence 18. File the resume, *then after you stamp the schedule, sign the letter, and last,* copy the diploma.

Sequence 19. Copy the agreement, *then fax the questionnaire, after you scan the budget form,* and last, print the memo.

Sequence 20. Use the spray, *then drink the syrup, before you administer nasal drops,* and last, take the tablet.

Sequence 21. Scan the diploma, *then after you punch the report*, *fax the budget form,* and last, file the checklist.

Sequence 22. Administer eye drops, *then before you take the capsule, drink the syrup,* and last, use the spray.

Sequence 23. Fax the certificate, *then scan the order form, before you copy the memo,* and last, print the proposal.

Sequence 24. Stamp the lease, *then print the manual, after you file the transcript, and last*, scan the certificate.

Sequence 25. Take the tablet, *then before you administer ear drops, drink the syrup,* and last, use the spray.

Sequence 26. Stamp the checklist, *then after you sign the claim, file the memo,* and last, copy the manual.

Sequence 27. Sign the letter, *then print the transcript, after you stamp the questionnaire,* and last, copy the paycheque.

Sequence 28. Punch the receipt, *then scan the manual, before you file the agreement,* and last, fax the pay stub.

Sequence 29. Take the capsule, *then after you administer nasal drops, use the spray,* and last, drink the syrup.

Sequence 30. Copy the lease, *then before you punch the budget form, sign the schedule,* and last, fax the order form.

Sequence 31. Print the contract, *then file the letter, before you stamp the record, and last,* scan the paycheque.

Sequence 32. Punch the invoice, *then stamp the contract, after you sign the licence,* and last, file the resume.

Sequence 33. Drink the syrup, *then before you use the spray, take the tablet,* and last, administer eye drops.

Sequence 34. Stamp the questionnaire, *then after you scan the complaint form, fax the report,* and last, sign the notice.

Sequence 35. Fax the certificate, *then copy the declaration, after you sign the proposal,* and last, punch the agreement.

Sequence 36. Print the receipt, *then file the invoice, before you punch the pay stub,* and last, copy the resume.

Sequence 37. Administer nasal drops, *then after you take the capsule, drink the syrup,* and last, use the spray.

Sequence 38. Stamp the checklist, *then before you scan the proposal, print the complaint form,* and last, file the receipt.

Sequence 39. File the statement, *then punch the notice, before you scan the receipt,* and last, fax the letter.

Sequence 40. Scan the licence, *then print the transcript, after you stamp the statement,* and last, file the diploma.

Appendix G

Two samples of the enactment slides in Experiment 3