10-2014

Reexamining Linguistic Relativity: What Adult Bilinguals Can Teach Us About Culture, Language, And Cognition

Natalya Petroff
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!
Follow this and additional works at: https://academicworks.cuny.edu/gc_etds
Part of the Educational Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/371

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
REEXAMINING LINGUISTIC RELATIVITY: WHAT ADULT BILINGUALS CAN TEACH US ABOUT CULTURE, LANGUAGE, AND COGNITION

by

NATALYA PETROFF

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2014
This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Bruce D. Homer

Date

Chair of Examining Committee

Dr. Alpana Bhattacharya

Date

Executive Officer

Dr. Jens Brockmeier

Dr. Howard T. Everson

Dr. Gigliana Melzi

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

REEXAMINING LINGUISTIC RELATIVITY: WHAT ADULT BILINGUALS CAN TEACH US ABOUT CULTURE, LANGUAGE, AND COGNITION

By

Natalya Petroff

Advisor: Professor Bruce D. Homer

Extending Whorf’s popular notion of linguistic relativity (LR) to bilingual contexts, one would argue that a speaker’s first language (L1) influences her thinking and behavior under second language (L2) conditions. According to one interpretation of LR, *inter-language relativity*, L1 instills in its speakers habitual ways of thinking and thus influences their perception and categorization in L2 contexts. Under *intra-speaker relativity*, bilinguals follow either L1 or L2 patterns of performance, depending on L2 proficiency. Finally, according to usage-based accounts of language, there is no qualitative difference between mono- and bilingual speakers, and a bilingual's performance under L2 conditions is best viewed in terms of their ongoing engagement with L2.

To investigate how much each interpretation contributes to our understanding of cognition, language, and culture, two studies were conducted with a sample of 45 adult Russian-English bilinguals. Each study was based on a popular research paradigm and tested all three interpretations of LR for their explanatory value. Study one utilized a one-word association task conducted in both languages, a common way to examine the conceptual organization of the bilingual lexicon. Study two utilized a different kind of association task to investigate influences of L1 (grammatical gender) under L2 conditions. In both studies, there was no evidence in support of either inter-language or intra-speaker relativity. There was evidence in support of
usage-based accounts of language: bilinguals' use of informal English appeared to moderate their performance under L2 conditions.

*Keywords*: linguistic relativity; language and thought; grammatical gender; bilingualism; cultural psychology
Acknowledgements

I would like to express my gratitude for the support and guidance to the many people, both inside and outside academia, who made the completion of my dissertation possible:

First of all, I am deeply grateful to my advisor, Dr. Bruce Homer, for helping me navigate matters as diverse as placing my research within a larger theoretical context to figuring out a novel dissertation format; for being very generous with his time; and for imparting his belief in my success to me. Next, I am very appreciative of the support of Dr. Howard Everson: his technical expertise in data analysis and passion for precision have taught me a new appreciation of the role of measurement in research. Also, I would like to express my gratitude to Dr. Jens Brockmeier, who has helped me recognize that ambitious theoretical claims are part and parcel of empirical enquiry. I would also like to thank my readers, Dr. Alpana Bhattacharya and Dr. Gigliana Melzi, who graciously gave their time to support and assist me in my oral defense.

I am eternally thankful to my family who have sustained me in this long process: to my mother Lydia Petrovna for being an inspirational role model and my devoted champion. To my husband Yuri for his unconditional love and unwavering support through the many ups and downs of this long adventure. To my son Dmitry for generously giving me his time and support from day one and for never failing to cheer me on along the way.

Last but not least, I would like to acknowledge the many members of the Russian community who volunteered to participate in my study: their time and effort stand as proof of their confidence in me and the research community in general.

All in all, I feel blessed to have received so much support and inspiration from so many people. Thank you all!
Table of Contents

INTRODUCTION .................................................................................................................................................. 1

CHAPTER ONE: STUDY ONE .......................................................................................................................... 5

  Linguistic Relativity: From Inter-language to Intra-speaker ................................................................. 6
  Language experience: The missing link? ............................................................................................... 10
  Study Description .................................................................................................................................. 13

Method ........................................................................................................................................................... 15

  Participants ................................................................................................................................................ 15
  Procedure and Instrumentation ............................................................................................................ 16
  Measures .................................................................................................................................................. 20

Results ........................................................................................................................................................... 23

  Descriptive Findings ........................................................................................................................... 23
  Preliminary Findings ............................................................................................................................. 26
  Testing Inter-language and Intra-speaker Theories .............................................................................. 28
  General Cognitive Mechanism Hypothesis ............................................................................................ 29

Discussion ..................................................................................................................................................... 33

CHAPTER TWO: STUDY TWO......................................................................................................................... 41

Introduction .................................................................................................................................................. 42

  Study Description .................................................................................................................................. 47

Method ........................................................................................................................................................... 48

  Participants ................................................................................................................................................ 48
  Procedure and Instrumentation ............................................................................................................ 49
  Measures .................................................................................................................................................. 52

Results ........................................................................................................................................................... 54

  Descriptive Findings ........................................................................................................................... 54
  Preliminary Findings ............................................................................................................................. 55
  Testing Inter-language Theory .............................................................................................................. 56
  Testing Intra-speaker Theory .................................................................................................................. 58
  Exploring “Reverse” Masculinity Bias .................................................................................................... 59

Discussion ..................................................................................................................................................... 64
CONCLUSION .................................................................................................................. 71
APPENDICES .................................................................................................................. 74
Appendix A: Screening Form ......................................................................................... 75
Appendix B: Consent Form ............................................................................................ 76
Appendix D: Language History Questionnaire ............................................................. 80
Appendix E: Cue Words .................................................................................................. 87
Appendix F: Adjective Rating Form ............................................................................. 88
Appendix G: List of Pooled Adjectives ........................................................................ 90
REFERENCES ............................................................................................................... 91
List of Tables

Table 1 Participant Characteristics ........................................................................................................... 23
Table 2 Means (M), Standard Deviations (SD), Ranges, and t-test Results for the Demographic Variables by L2 Fluency Group ........................................................................................................... 25
Table 3 Means (M) and Standard Deviations (SD) of Predictors and Outcome Variables .............. 26
Table 4 Correlations between Predictors and Outcomes as a Function of L2 Fluency ................. 27
Table 5 Results of t-tests Comparing Low and High L2 Fluency Bilinguals on L1-L2 Match and L2 Commonality ................................................................................................................................ 29
Table 6 Summary of Hierarchical Regression Analysis for Variables Predicting Consistency among Late-onset Bilinguals (N = 45) ........................................................................................................... 31
Table 7 Means (M), Standard Deviations (SD), and t-test Results for the Demographic and Language Use Variables by L2 Content ........................................................................................................... 55
Table 8 Means (M), Standard Deviations (SD), and t-test Results for the Demographic and Language Use Variables by L2 Content ........................................................................................................... 56
Table 9 Estimated Marginal Means (M) for the “Reverse” Masculinity Bias Effect ................. 63
List of Figures

Figure 1 Interaction of L2 Fluency and Informal English to Predict Consistency among Late-onset Bilinguals ................................................................. 32

Figure 2 "Reverse" Masculinity Bias ........................................................................ 57
INTRODUCTION

Whorf’s (1956) claim that language was implicated in thinking and behavior in non-trivial ways – or linguistic relativity (LR) – was indeed revolutionary: it paved the way for the study of language as both a linguistic and psychological phenomenon\(^1\). The idea has produced many provocative lines of research bringing together psycholinguists, applied linguists, and cognitive and developmental psychologists (e.g., Fausey & Matlock, 2011; Gentner & Goldin-Meadow, 2003; Gomila, 2012). For example, cognitive psychologists (Boroditsky & Schmidt, 2000) have looked at grammatical gender as a factor influencing categorization among speakers of different languages. The diversity of the field is stimulating: it invites more research to explore LR, building on and challenging the existing studies. To that end, a series of two studies was conducted and provided the content of the present dissertation.

To organize this growing body of empirical findings on the impact of language on cognition, Reines and Prinz (2009) have proposed several interpretations of LR, all of which share the same starting point: a speaker’s first language (L1) sets her along a certain cognitive trajectory, effects of which persist even under second language (L2) conditions. To this inter-language reading of LR, researchers studying monolingual and bilingual speakers have offered a modified – *intra-speaker* – version: Kousta, Vinson, and Vigliocco (2008) have suggested that bilinguals tend to follow either L1 or L2 patterns depending on language context. Both interpretations of LR, inter-language and intra-speaker, view language as representational in nature, with its primary function being to reflect reality and present us with an accurate picture of the world.

\(^1\) Prior to Whorf, 19\(^{th}\) century thinkers Peirce, Saussure, and Humboldt had considered language and thinking as interrelated phenomena. Their subsequent influence has generally been limited to the field of semiotics.
This view of language as a depository of knowledge – stable, universal, and a-temporal – is challenged and its limitations become apparent once we expand the scope of our inquiry to speakers of different languages, ages, and cultural backgrounds (Wierzbicka, 2014). In an attempt to overcome these difficulties, I decided to turn to alternative, non-representational approaches to language and LR offered by usage-based grammar, research on second language acquisition (SLA), and cultural psychology. They all share the view of language as grounded in meaningful social practice, so that its main function is communicative and social (e.g., Croft, 2013; Deacon, 1997; Tomasello, 2008). Accordingly, language has an important impact on speakers through their ongoing engagement, contact, and action with other speakers. In this view, the only true universals are needs and goals of any human community (Tomasello, 2003). For example, Fox and Thompson (2010) have demonstrated how English syntax is part and parcel of live English conversation: when responding to a clarifying question, speakers use a clausal (sentence-like) response to signal a potential break in communication; conversely, a phrasal (one-two word) answer points to mutual understanding and smooth communication flow.

The view of language as serving crucial social needs of a community of speakers underscores the importance of experiential (non-linguistic) factors such as quantity and quality of L2 experience and use. As Moyer (2011) has shown, L2 proficiency among late-onset bilinguals is predicted by length of residence and L2 use with friends, who are native speaker of English, or quantity and quality of L2 experience respectively. Additionally, SLA researchers studying bilingual adults have long acknowledged that the field tends to rely on populations found in college settings (Larotta, 2010; Moss & Ross-Feldman, 2003), which result in findings that may not generalize to older, non-academic samples. In response to these developments, I decided to study a community-based sample of older bilingual speakers, anticipating that a focus on
bilingual adults, who had acquired L2 post-puberty, would have the potential of further explaining how language and sociocultural factors interact in bilingual development.

In terms of study design, applying a three-part theoretical backdrop offers certain advantages because each reading of LR has a well-developed focus and, thus, generates specific predictions about the role of language as a destiny, a repertoire of responses, or (trans)formative experience. To test all three frameworks, I used two separate word-association tasks. In my choice of task, I was guided by what is customary in the field of LR: word association tasks have often been used to study language and cognition in bilingual samples (Meara, 2009). The format is flexible enough to accommodate virtually any combination of languages. By modifying stimulus selection and instructions, a researcher can test various linguistic and cognitive factors, such as whether and how L1 vocabulary and L1 grammar might affect L2 conceptual associations in bilinguals. Also, given its long history and popularity among various disciplines, I thought it would be important to explore ways of enhancing this task further through the combination of formal and experiential measures in my research.\(^2\)

Finally, given the variety of disciplines that have contributed to the field of LR, a two-study format was an opportunity to consider several research threads within the context of the same research project. Also, since there has been a dearth of non-representational theory in the field, running two independent tasks allowed me to examine whether a non-representational approach offers any advantages compared to traditional, representational approaches. A trend across two tasks should be an indication of which approach holds more potential. Last but not least, expanding my sample to include older community-based participants should contribute to

---

\(^2\) Some researchers have been questioning this task’s utility in various contexts, hence, the telling title of one research study, “Assessing proficiency through word associations: is there still hope?” (Wolter, 2002). For a very comprehensive history of the subject, see Meara (2009).
the field of second language acquisition and help us explain how adults learn L2 as well as bring about new educational opportunities.
CHAPTER ONE: STUDY ONE

LINGUISTIC RELATIVITY THROUGH THE LENS OF LANGUAGE EXPERIENCE:
AN EXPLORATORY STUDY OF COGNITIVE EFFECTS OF BILINGUALISM AMONG
LATE-ONSET BILINGUALS
Linguistic Relativity: From Inter-language to Intra-speaker

It has been more than a half of a century since Benjamin Whorf (1956) put forward the assertion that the language we speak shapes our thought and action. His claim, known as linguistic relativity (LR), continues to generate a lot of interest across different fields, including cognitive psychology, applied linguistics, and philosophy of language, among others.

As more and more empirical findings become available, there seems to be less agreement among researchers about the best way to interpret these findings in a context of LR to advance our research, instruction, and policy. Reines and Prinz (2009) have offered a framework to organize the available data and guide future research. According to their classification, some researchers seem to interpret LR as Ontological Whorfianism, or the view that a speaker’s first language instills in her certain habitual patterns of thought that later become a lens through which she perceives and organizes her experience along certain concepts. On the other hand, Kousta, Vinson, and Vigliocco (2008) have suggested that we can now speak of two main flavors of LR: inter-language and intra-speaker. Ontological Whorfianism would be an example of the former. On the other hand, many studies with bilingual samples tend to rely on the latter interpretation of LR and find that bilinguals switch between languages and follow either L1 or L2 patterns depending on task context (e.g., language of instruction).

Research on vocabulary development among adult L2 speakers seems to provide empirical support to the Ontological Whorfianism view. For example, word recognition and translation studies with bilinguals have found that a speaker’s first language (L1) influences their second language (L2) vocabulary development supposedly because bilingual speakers access their conceptual system via the L1 lexicon. In other words, a bilingual’s lexical performance in an L2 context seems to depend on how fast and accurate they could map an L2 onto their L1
Early studies were guided by two models: word-to-word association theory (also known as lexical access), which proposes that an L2 is built on the L1 lexicon; and concept mediation whereby both L1 and L2 tap into the same conceptual storage (for a review see Francis, 1999). Later, Kroll and Stewart (1994) proposed a middle ground solution, the Revised Hierarchical Model (RHM). With the RHM, they introduced a developmental element to account for differentiated performance among novice and advanced second language learners: as a speaker becomes more proficient in an L2, she is able to access the conceptual system directly, bypassing the L1 lexicon. After 20 years since its introduction, the model is still influential across populations and tasks (e.g., Harris, Berko Gleason, & Aycicegi, 2006). Many word-association studies with bilinguals continue to utilize it and examine their findings in light of the interaction between L1 and L2 lexicons (e.g., Zareva, 2010).

With more research on novice and advanced bilinguals, though, the view of linguistic relativity as an inter-language phenomenon has been challenged. There has been a growing recognition that, perhaps, linguistic relativity is an intra-speaker phenomenon (Kousta et al., 2008). That is, once we compare different groups of bilinguals, low and high proficiency, there is little evidence of direct influences of L1. For example, Zareva, Schwanenflugel, and Nikolova (2005) compared two groups of Bulgarian-English bilinguals, intermediate and proficient as based on a standardized test of English (TOEFL), to a group of native English speakers (bilinguals were based in Bulgaria; native English speakers were US-based). All were assessed for breadth and depth of vocabulary. Participants associated to a list of 73 words and then received a score of native-like commonality of associations. Proficient bilinguals and native

---

3 Although, there has been some shift toward alternative views (e.g., De Groot’s Distributed Feature Model as elaborated by Brysbaert and Duyck, 2010).
speakers scored similarly on vocabulary measures but not on native-like commonality. In addition, both proficient bilinguals and native speakers showed stable patterns of word associations in contrast to intermediate bilinguals, whose word associations were more varied. It was also found that as bilinguals gained more proficiency, their associations became more native-like. The researchers offered that ultimately non-native speakers could never be native-like in their associations; rather, word association tasks can help us gauge a speaker’s L2 proficiency based on native-like commonality of associations. It is important to note that in the Zareva et al. study, there was no indication whether proficient bilinguals spoke with (near) native proficiency. Moreover, for their testing selection, the researchers took care to include words that were not part of everyday discourse and thus would have required extensive reading experience.

In a follow-up study, Zareva (2010) compared three groups of speakers, monolingual English and Bulgarian speakers and a group of Bulgarian-English bilinguals. All participants provided up to three written associations to a list of 36 words of different frequencies; bilinguals completed an English version of the list. Advanced bilinguals were more like English monolinguals in their associations than intermediate bilinguals. But overall, bilinguals’ associations were unlike that of monolinguals in either language in respect to the number of common associations: the two monolingual groups did not differ in their ratio of common v. idiosyncratic associations; in the bilingual group, however, there were significantly more idiosyncratic responses, 66% idiosyncratic v. 32% common. The researcher did not specify if there was a difference in the proportion of common v. idiosyncratic responses among advanced and intermediate bilinguals.

In a related line of research, Fitzpatrick (2006) has found that native and non-native English speakers tend to prefer different types of associations. Both groups were given a list of
60 words and asked to produce one-word associations. Stimulus words were selected from a list of college-level English (Coxhead, 2000), with the first two thousand high frequency words removed from the list. Non-native speakers also took a written test of receptive English as a measure of English proficiency. Word associations were then grouped into 16 categories (form-, meaning-, or position-based; erratic; and others) based on interviews with participants.

Fitzpatrick found distinctive patterns of association for each group of speakers. Native English speakers showed a preference for meaning-based associations (e.g., ‘confirmed > definite’), with position-based associations (collocations) coming in second (e.g., ‘visual > aid’). Non-native English speakers produced the same amount of meaning-based pairs as the native group but showed no preference for collocations over form-based. That is, native speakers produced twice as many collocation responses compared to the non-native group. Finally, in contrast to Zareva and colleagues’ research, Fitzpatrick found no correlation between English proficiency and non-native patterns of responses. Fitzpatrick’s findings also contradict some early studies that found that with a rise in proficiency, bilinguals tend to use more meaning-based and collocations and fewer form-based pairs (e.g., Meara, 1983).

In a follow-up study, Fitzpatrick (2007) clarified whether it makes sense to speak of native “norms” at all. She administered two versions of a written word association task (100 words) to a group of 30 native English speakers twice. The researcher used the same Coxhead list to avoid “stereotypical” responses. This time, classification of responses was done by a native English speaker. She found that meaning-based associations (or synonyms as in ‘confirmed > definite’) and collocations were equally present; more specifically, consecutive collocations (or cue-target as in ‘visual > aid’) were the top choice (similar to Nissen & Hendriksen, 2006). This finding challenges the traditional representational view, which is still
very common in the field, i.e., “The EL1 lexicon is paradigmatically dominated” (Zareva, 2010, p. 18). Fitzpatrick also found a large degree of between-speaker variation (e.g., some speakers produced collocations only). Overall, however, participants showed stable patterns of associations over time. As a result, Fitzpatrick disputed the idea of native speaker norms as a means of assessing bilingual development. According to her, the idea that there is a progressive move from syntagmatic (collocations) to paradigmatic (meaning-based) associations originated in research on language development: some studies have found that as children mature, they move away from syntagmatic to paradigmatic associations. Likewise, in bilingual research it is now expected that L2 speakers should follow the same “developmental” trajectory. Having found no evidence of native “norms” in her research with native English speakers, Fitzpatrick disputes that word-association patterns have anything to do with maturity or L2 development. She instead suggests that with more proficiency bilinguals should show the same associative style in either language.

Language experience: The missing link?

As the previous discussion has demonstrated, studies with bilingual samples have moved away from viewing bilinguals as either novices or proficient L2 speakers, toward a more nuanced approach. More studies have been done with intermediate and more advanced bilinguals; bilingual proficiency has been measured by standardized written tests instead of relying on demographics such as age or length of residence. Still, little has been done to move beyond traditional measures of L2 competence, such as off-line, written tests, and to explore which experiential factors contribute to bilingual proficiency in adult speakers and how differences in L2 experience contribute to cognitive outcomes in adult bilinguals.

4 To account for the shift, some have suggested formal schooling (Cole, 1990; Nelson, 1977) and literacy (Cronin, 2002). Research with bilingual children has provided evidence of the shift taking place along the same timeline as in monolingual children (Shen, McGregor, & Marian, 2006).
Often, L2 experience is studied under the blanket term of “age of onset”. This undifferentiated approach should not be surprising given that for a long time bilingual development was viewed as an extension of language development subject to *maturational constraints*: that is, language development is closely aligned with and driven by biological and neural maturation, so that once particular neurological milestones are met, certain linguistic developments become less likely. Following this line of reasoning, L2 experience was often reduced to age so that bilingual speakers were studied as different age groups, e.g. early- v. late-onset. For example, one influential study (Johnson & Newport, 1989) found that among early-onset Chinese-English and Korean-English bilinguals (age of onset < 16) there was a negative correlation between age and L2 proficiency. On the other hand, among late-onset bilinguals there was no relation between age and L2 proficiency. The researchers interpreted these findings as evidence of maturational constraints at work: age was a predictor of L2 proficiency among early-onset bilinguals; yet age was not a predictor among late-onset bilinguals, purportedly because their L2 acquisition took place after they had already passed their maturation period.

These findings were later challenged as well as the whole maturational paradigm. The challenge came from researchers who viewed L2 development as predicated on language experience and everyday practice. For example, Birdsong and Molis (2001) replicated the Johnson and Newport study and found evidence against maturational accounts of bilingual competence: there was no relationship between age and L2 proficiency among early arrivals (age of onset < 16); there was a negative relationship between age and L2 proficiency among late arrivals. Moreover, among late arrivals, amount of L2 use was positively correlated with L2 proficiency which prompted the authors to note that “[the] results suggest a nontrivial role for practice in determining the L2 acquisition end state” (p. 247).
In a recent study of bilingual proficiency and language experience, Moyer (2011) examined 42 bilinguals of various linguistic backgrounds and various levels of English (English/L2 proficiency was operationalized as accent). She found that several experience-related factors correlated with accent, including length of residence, age of onset (first exposure to either formal or informal English), weekly use of L1, and weekly informal use of L2 (with friends, native English speakers). In contrast, weekly formal use of L2 (in academic settings) did not correlate with accent. A follow-up multiple regression test showed that once weekly informal use of L2 is entered, age of onset becomes non-significant, while length of residence and weekly informal use of L2 remain significant. Overall, Moyer’s study provides additional evidence against maturational accounts of bilingual competence: age turned out to be less important than length of residence and weekly informal L2 use, or quantity and quality of language experience, respectively.

Summarizing the literature, Marinova-Todd et al. (2000) have noted that there is little if any conclusive evidence of bilingual competence as an age-related capacity. Rather, L2 development depends on what kind of language experience bilinguals are exposed to on a regular, continuous basis. Until researchers recognize the importance of language experience, there would be more misinterpretation, misattribution and mis-emphasis in the field and as a result less progress.

In a related line of studies on language and its role in human development, researchers have emphasized the need to recognize that, first and foremost, language plays a very important communicative and social role; thus, we need to move away from viewing language in strictly linguistic terms, a-temporal and universal. From this perspective, what makes language universal are the common communicative and social goals, which all human communities need to
accomplish. In other words, the emphasis is not on describing various linguistic structures such as vocabulary or syntax for their own sake, but on figuring out how language works in terms of its advancing communicative goals (e.g., Croft, 2013; Port, 2013; Tomasello, 2008) and the underlying cognitive processes that organize speech, such as chunking or analogy (e.g., Beckner et al., 2009). For example, Ford and Thompson (2010) have shown how different syntactic structures serve different roles in English discourse: speakers rely on phrasal (one-two word) responses to signal they are on the same page with their interlocutor; in contrast, they use clausal (sentence-like) responses to indicate communication breakdown. In other words, non-linguistic, experiential and contextual factors shape speaker behavior. Finally, ignoring non-linguistic factors seems to hinder traditional psycholinguistic research: Fitzpatrick (2007), a linguistically-oriented researcher who studies word associations among bilinguals, has suggested that we need to identify “non-linguistic factors affecting association behavior” (p. 327) if we are to understand the role of language in bilingual development.

**Study Description**

The present study sought to investigate several interpretations of linguistic relativity in a bilingual context and explore how much each interpretation contributes to our understanding of how language, cognition, and culture interact. Following a well-established tradition of using word associations to study L1 influences on L2 among bilinguals, a word-association task was conducted in two languages. The task, coupled with standardized measures of L2 proficiency and self-reports of uses of L2, provided data to test the following three predictions, each informed by a specific interpretation of linguistic relativity:

**Inter-Language Theory.** In its inter-language form, linguistic relativity proposes that as bilingual speakers move between languages, their first language continues to influence their
cognitive performance under second-language conditions. According to Ontological Whorfianism, speaking a language leads to the creation of categories that the speaker then uses to organize their experience in other languages. As a result, when we compare word associations in two languages, we should be able to find evidence of L1 influencing L2 associations since the conceptual storage is shared across the two languages (e.g., Fitzpatrick & Izura, 2011; for review, see Francis, 2005). In the present study, inter-language theory predicted that more proficient bilinguals should have more L1 and L2 association pairs that match compared to less fluent bilinguals because it is easier for the more proficient bilinguals to access their conceptual store.

**Intra-Speaker Theory.** According to this reading of linguistic relativity, a bilingual’s performance tends to change depending on the speaker’s current linguistic milieu so that a bilingual follows either L1 or L2 patterns in accordance with their L2 proficiency. Informed by bilingual studies, this view tries to account for different patterns found among more- and less-proficient bilinguals. Studies have shown that bilinguals may resemble either L1 or L2 group on word associations (Zareva, 2010; Zareva et al., 2005). In the present study, this framework predicted that more proficient bilinguals would produce L2 associations that are more similar to those of native English speakers, compared to less proficient bilinguals whose L2 associations should be less similar to native speakers’ associations.

**General Cognitive Mechanism (GCM) Theory.** In contrast to the other two readings, this framework does not view language as a competence that determines a speaker’s cognitive path. Nor does it believe that a bilingual speaker possesses two languages and switches between L1 and L2. Instead it suggests that after reaching a certain level of L2 proficiency, bilinguals demonstrate stable patterns of performance across language conditions. The issue of L2
proficiency then becomes important: traditional measures such as off-line written tests of L2 vocabulary may not be sensitive enough to distinguish between different levels of L2 proficiency and thus result in inconsistent findings. Additional measures such as on-line L2 fluency and self-reports of language experience, including uses of L2 in informal settings, may be necessary to provide for a more nuanced measure of bilingual proficiency. According to GCM theory, more proficient bilinguals should show more stable patterns of associations across language conditions compared to less proficient bilinguals, and it would be expected that L2 experience and use would also play a role in explaining the difference in performance between the two groups of bilinguals.

To investigate the three interpretations of linguistic relativity, a group of late-onset Russian-English bilinguals performed a word-association task in both languages, filled out a language experience survey and completed a standardized test of spoken English. Their task performance was scored for three outcomes: between-language conceptual consistency, native-likeness of responses, and consistency of associations. Measures of L2 proficiency and L2 use were also taken. The outcome measures were then examined for evidence for and against the three theories.

Method

Participants

Forty-five late-onset Russian-English bilinguals took part in the research ($M_{age} = 37.56$ years, $SD = 9.44$; 10 male). To participate in the study, they had to be functionally fluent in

---

5 The sample size was determined based on the results of a pilot study, which produced a Cohen’s $d$ of .8. A power analysis conducted with an open-source statistical tool (Soper, 2014) demonstrated that at least 42 participants were needed to obtain an anticipated effect size of .8, with a desired power of .8 and alpha of .05 for one-tailed $t$-tests and regression analysis with two predictors. The final sample of 45 was recruited to deal with unanticipated adverse events that could have resulted in attrition and loss of data. In the end, all 45 participants qualified for and concluded the study successfully.
Russian (L1) and English (L2) and use both languages on a regular basis across settings (e.g., home, work, with friends). To qualify as late-onset Russian-English bilinguals, they were expected to have spoken Russian as their first language; had not had prior significant exposure to English as their ambient language before coming to the US; had arrived to the US as young adults, after graduating high school in their native country (some participants were from former Soviet republics). In addition, they had to have never been diagnosed with a language disorder or as learning disabled. All participants were screened over the phone to ensure that they met the study’s selection criteria (Appendix A). The screening procedure was conducted in Russian to assess their command of spoken Russian.

Subject recruitment utilized flyers, word of mouth, and postings on social networks. Participants received gift cards redeemable at local book stores ($20) and a copy of their English proficiency report, based on a standardized test of spoken English, which was part of the study.

Procedure and Instrumentation

All participants were tested individually in a psychology lab by the author. After a brief introduction in Russian to induce a bilingual mode and activate Russian as their base language (Grosjean, 2001)⁶, participants completed several tasks and language measures in the following order: a one-word association task in Russian; then English was introduced as the base language as participants carried out a language experience survey in English; they then completed a one-  

---

⁶ It is common in bilingual research to use a bilingual speaker’s native language (L1) for part of the study and thus activate both languages and engage a bilingual mode (e.g., Marian & Spivey, 2003). Known as language mode theory, this view purports that bilingual performance is subject to monolingual and bilingual process activation, based on linguistic and contextual factors (Grosjean, 2001). Language mode theory suggests that there is a continuum of monolingual/bilingual activation, with either language, L1 or L2, capable of being the base language depending on multiple factors, such as bilingual proficiency and context. It is generally accepted that bilingual participants may be put into a bilingual mode by any of the following factors: if they know that they are participating in a study on bilingualism, if they are aware that the experimenter is bilingual, if they see or hear stimuli from both languages (Grosjean, 2001). In the present study, all three factors were at play: the bilinguals knew that this was a bilingual study; that the author was bilingual; and they had to use both languages. Moreover, it would be safe to say that in the first part of the study, Russian was the base language of the bilinguals: they used Russian to communicate with the author when they first entered the lab (following the author’s lead), it was then used as language of instruction, and finally, as stimuli in the first Russian version of the word-association task.
A word association task in English, an L1 grammar association task in English and, finally, a standardized test of spoken English (the data of the L1 grammar association task is not part of the current study). All tasks were untimed. On average, completion of the entire protocol took about one hour.

**One-Word Association Task.** The word-association task consisted of a series of 35 stimulus words (see Cue Words below). The series was composed of Russian nouns and adverbs with a single or dominant translation in English. Participants were asked to read aloud the stimulus word as presented on a computer’s screen, say a one-word association and then move on to the next screen by mouse click. The task was not timed but participants were asked to work at a brisk pace. After completing a series of four trial screens, they completed the 35-word series. Stimulus words were given in random order. Responses were recorded in two formats: audio format for subsequent transcription and text format, allowing for time-stamping the data to measure response time (the response time data is not part of the current study). An English version of the task, based on the same items translated into English, was given later in the session after a participant had completed a language experience survey in English. All participants thus had a chance to switch from Russian as their base language to English before taking the English version of the task.

**Cue Words.** Thirty five cue words were selected for their frequency and translation properties (Appendix C). In this line of research, it is common to select words of different frequency (high, medium, low) so that study participants may be assessed for L2 proficiency based on their L2 vocabulary (e.g., Zareva et al., 2005). However, since the focus of the present study was on interaction between L1 and L2 based on online performance, only high-frequency words were selected to control for possible differences in vocabulary development. Additionally,
care was taken to ensure that all selected cues would also be present in a database of native speakers’ responses, which is based on high-frequency English words (Nelson, McEvoy, & Shreiber, 1998).

It must be noted that there is an inherent tension in this kind of association research with bilinguals: researchers strive to make their cue selection representative yet they have to eliminate certain types of cues to avoid stereotypical responses and translational ambiguity (multiple translations). For example, some researchers have asserted that high frequency concrete nouns tend to map directly from one language to another and thus result in stereotypical responses and associations (e.g., Kroll & Stewart, 1994). However, recent studies with high frequency concrete words have challenged this view (e.g., Pavlenko & Malt, 2011).

In addition, researchers have to guard against word classes that result in multiple translations. For example, Prior, MacWhinney, and Kroll (2007) have found that compared to English nouns, English verbs produce a greater number of translations in Spanish. Also, it is rather common for English words to belong to more than one word class which also contributes to multiple translations (e.g., the word “finger” could be either a noun or a verb; the word “frequent” could be either a verb or an adjective).

To ensure that the list is representative, without being prone to translation ambiguity, several steps were taken. First, a popular list of English words, known as the Toronto Word Pool7 (Friendly, Franklin, Hoffman, & Rubin, 1982) was checked for words that were exclusively nouns or both nouns and verbs (an equal number of concrete and abstract words were selected); next, seven adjectives were selected. All 35 words were checked against the Nelson et

---

7 The Toronto Word Pool is used in many areas of research, including psycholinguistics, memory studies as well as studies with clinical populations.
al. database (Nelson et al., 1998) to ensure that semantic network information would be available for them.

As is customary in word association research (e.g., Prior et al., 2007), Russian words and their English translations were then checked for translation equivalency based on translations provided by four Russian-English bilinguals (who were not part of the study). In addition, all Russian cues were checked against the Oxford English-Russian Bilingual Dictionary (Oxford Language Dictionaries Online, n.d.) to verify that they all had only one or dominant translation in English. Finally, five training items were selected, with the same ratio of parts of speech and concreteness measures as the testing list.

**Language Experience Survey.** The Language Experience Survey was adapted from the Li, Sepanski, and Zhao (2006) questionnaire (see Appendix D). The survey included sections on demographics, self-ratings of English and Russian proficiency across modalities, estimates of language use in different social settings and for different purposes, and others. Only the sections on demographics, English language use, and Russian language use were utilized in the current study (for detailed description of language use data, see Measures).

**Test of Spoken English.** To assess for L2 proficiency, participants completed a standardized test of spoken English (http://www.versanttest.com): the Versant Test of Spoken English is administered over the phone and instantly produces several measures. Designed to assess speakers’ facility with spoken English, the test is commonly used by academic and applied researchers working with bilinguals (e.g., Bradlow et al., 2011; Douglas, 2013). The test provides one Overall score and four sub-scores based on four subscales that cover two broad areas, content and control (subscale range: 20 – 80). The former refers to the ability of a speaker to understand spoken English and thus taps into their grasp of English grammar and syntax.
(Sentence Mastery and Vocabulary). The latter refers to the ability of a speaker to express themselves clearly in spoken English. There are two subscales to measure a speaker’s command of spoken English: Fluency and Pronunciation. Fluency examines a speaker’s rate of speech and rhythm while Pronunciation reflects a speaker’s ability to produce English phonemes in a native-like manner in phrasal contexts. The test has strong reliability and validity: the split-half reliability coefficients for the two sub-scales range from .97 to .99. Also, validation studies have shown that Versant test scores highly correlate with other measures of English proficiency such as the TOEFL (Bernstein, Van Moere, & Cheng, 2010). Given that both subscales assess related areas (control) and the two subscales highly correlate with each other ($r = .80$), the two sub-scores were combined to produce one score as a measure of their on-line spoken proficiency in English (see L2 Fluency under Measures below).

**Measures**

Participants were given a randomized sequence of 35 cue words (see Appendix C), first in Russian (L1) and then English (L2), and produced two sets of associations. The L1 and L2 pairs were then examined for L1-L2 conceptual/lexical similarity for each participant (see L1-L2 Conceptual Mapping Score below). Next, L2 target words were compared against a list of targets produced by native English speakers (see L2 Commonality Score below). Finally, each participant’s L1 and L2 associations were scored for integration across L1 and L2 (see consistency Index below).

**L1-L2 Conceptual Mapping Score.** To gauge a speaker’s ability to access the conceptual store equally well from either language, a measure of conceptual/lexical consistency across the two languages was calculated. First, L1 targets were translated into English and then compared to the corresponding L2 targets. Each L1-L2 target match was scored as 1; not-
matching L1 and L2 targets as 0. Each participant received an *L1-to-L2 Conceptual Mapping Score* which was calculated as the Total Number of L1-L2 (conceptually/lexically) Matching Pairs (range: 0 – 35; e.g., “0” – no matches, “35” – all 35 pairs match). A higher L1-L2 conceptual mapping score should indicate better access to the (shared) conceptual storage.

**L2 Commonality Score.** To measure native-like associative behavior or how closely bilinguals follow native English norms, an *L2 Commonality Score* was calculated. L2 cue-target pairs were assessed for similarity with native English norms as presented in the Nelson et al. (1998) word association database. The database provides norms for 5,019 English words of different parts of speech, normed on groups of 138 to 198 participants of college age. Each word (cue) comes with a network of associations; networks may have as few associations (targets) as 1 or as many as 30. For each target, an index of forward strength or cue-to-target strength is provided; it is calculated as Number of Participants Who Produced Cue-Target Pair over Group Size. The measure, in other words, refers to the probability that any given target will be produced in response to its cue. All L2 targets that were part of the Nelson database received a frequency score based on their frequency score in the database; for each participant an L2 commonality score was calculated as the average of (native) frequency scores of their L2 targets (range: 0 – 1; 0 – no similarity with native English speakers; 1 – very similar to native English speaker⁸). A higher L2 commonality score should indicate that the bilingual closely follows L2 patterns of associations.

**Consistency Index.** Instead of using formal linguistic categories that may not transfer well across languages or comparing bilingual performance to native samples⁹, an alternative

---

⁸ The maximum value (1) is virtually unattainable: the only way for a target to have a frequency of 1 would be if its cue had a network of 1: the Nelson database had one such pair: ‘left > right’, which was not part of the stimulus set.

⁹ For an alternative view, see Zareva (2010): instead of using native norms, we should use bilingual data to assess how well an English word is integrated into a bilingual speaker’s vocabulary.
measure of bilingual performance, *Consistency Index*, was created. The measure relies on on-line data and provides an estimate of L1-L2 integration as it captures individual association styles across language conditions. That is, each bilingual served as their own control and provided baseline when performing in their first language (similar to Fitzpatrick’s (2007) study). For each L1 cue, all targets were pooled together and then scored for their frequency. Next, each participant received an L1 bilingual commonality score as the average frequency of their L1 targets. A separate L2 bilingual commonality score was then calculated for each participant following the same procedure. Finally, each participant received a consistency index score calculated as the ratio of the two commonality scores, i.e., L1 commonality/L2 commonality; possible range: [0, +∞). A consistency index score that is close to 1 should indicate a high degree of L1-L2 integration.

**L2 Fluency.** To measure L2 proficiency, an L2 Fluency score was calculated for each participant as the sum of their fluency and pronunciation sub-scores on the Versant test (possible range: 80 - 160, where 20 is the lowest and 80 the highest score on each subscale)

**Language Experience.** Participants’ use of the English language was measured using four items from the language experience survey. The items asked participants to estimate in hours how much time they spend speaking English with their spouses, children, parents and relatives, and friends. Their responses were then added to produce a single informal English score for each participant. To that end, a composite variable, *Informal English*, was created after subjecting four Informal English sub-scores/items to a series of data reduction procedures. First, to normalize the data, all responses were re-coded as low (1), medium (2) or high (3). Then a composite informal English score for each participant was calculated as the sum of their four sub-scores in the informal English category (i.e., (i) spouse or partner, (ii) children, (iii) other
family members, and (iv) friends). The same kind of re-coding procedure was applied to the corresponding Russian categories to calculate informal Russian scores. The inclusion of this additional variable was necessary to explore whether informal English correlates with informal Russian. A significantly high level of correlation would be problematic as it would suggest that a speaker’s interpersonal style might be a factor. To examine whether informal English was representative of language experience and not indicative of interpersonal style, a correlation test was performed. The two variables, informal English and informal Russian were examined at the overall group level. The correlation coefficient was not significant at the p < .05 level, indicating that informal English was indeed a measure of L2 experience and not a measure of interpersonal style.

**Results**

**Descriptive Findings**

The mean and standard deviations of demographic variables are presented in Table 1.

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>37.56</td>
<td>9.44</td>
<td>24-58</td>
</tr>
<tr>
<td>AoA</td>
<td>26.32</td>
<td>6.27</td>
<td>18.5-53</td>
</tr>
<tr>
<td>LoR</td>
<td>11.23</td>
<td>6.98</td>
<td>1-24</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>35f, 10m</td>
</tr>
<tr>
<td>EDU</td>
<td>3.53</td>
<td>0.92</td>
<td>1-5</td>
</tr>
</tbody>
</table>

*Note. M = Mean; SD = Standard Deviation; Age = chronological age, in years; AoA = age of arrival in the United States; LoR = length of residence in the United States; EDU = educational attainment. EDU scores range from 0 (Did not Graduate High School) to 5 (Completed Post-Graduate Degree).*
As summarized in Table 1, on average, the mean age was just slightly over 37 years; the mean age of arrival was about 26. The participants had lived in the U.S. for about 11 years, and their average level of educational attainment was a bachelor’s degree.

To allow testing of inter-language theory and intra-speaker theory, two groups were created based on a median split: bilinguals with L2 fluency scores equal to or lower than 141 were classified as low L2 fluency bilinguals and those with scores equal to or greater than 143 as high L2 fluency. (The median split approach was chosen because the L2 fluency data was skewed: 13 out of 45 participants received the top score of 80 on each of the two subscales. Also, a two-group t-test produced a significant finding, indicating that the two groups were indeed different \( t(28.31) = -11.22, p < .001 \). Finally, taking the median value as the cut-off point resulted in two groups that were very close in size.) Next, a series of t-tests were performed to assess for any differences in demographic characteristics between low and high fluency bilinguals, see Table 2 below.
Table 2
*Means (M), Standard Deviations (SD), Ranges, and t-test Results for the Demographic Variables by L2 Fluency Group*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low&lt;sup&gt;a&lt;/sup&gt;</th>
<th>High&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
<td>M</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Age</td>
<td>39.36</td>
<td>10.48</td>
<td>25-58</td>
<td>35.83</td>
<td>8.19</td>
<td>24-52</td>
</tr>
<tr>
<td>AoA</td>
<td>28.1</td>
<td>7.32</td>
<td>20-53</td>
<td>24.13</td>
<td>4.16</td>
<td>18.50-33</td>
</tr>
<tr>
<td>LoR</td>
<td>10.75</td>
<td>8.19</td>
<td>1-24</td>
<td>11.70</td>
<td>5.75</td>
<td>2-21</td>
</tr>
<tr>
<td>Gender</td>
<td>16f, 6m</td>
<td></td>
<td></td>
<td>19f, 4m</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>EDU</td>
<td>3.45</td>
<td>1.06</td>
<td>1-5</td>
<td>3.61</td>
<td>0.78</td>
<td>2-5</td>
</tr>
</tbody>
</table>

*Note. Age = chronological age; AoA = age of arrival in the United States; LoR = length of residence in the United States; EDU = educational attainment. EDU scores range from 0 (Did not Graduate High School) to 5 (Completed Post-Graduate Degree).*

<sup>a</sup>n=22; <sup>b</sup>n=23.

* p < .05

As shown in Table 2, there were no significant differences between low and high L2 fluency participants with regard to their age, length of residence, and educational attainment. The two groups differed in their age of arrival in that low fluency bilinguals tended to come to the US at an older age (M = 28.1, SD = 7.32) compared to the high fluency group (M = 24.13, SD = 4.16); the difference of 4.48 (95% CI [0.93-8.04]) was statistically significant, t(43) = 2.54, p < .05. To explore any potential effect of this variable on outcome variables, it was tested for correlation with the outcome variables, see Preliminary Findings below.
Preliminary Findings

The means and standard deviations of all three outcome measures, along with two predictors, are summarized in Table 3 below\(^\text{10}\). The values are reported for the sample as a whole as well as by L2 fluency level.

Table 3
*Means (M) and Standard Deviations (SD) of Predictors and Outcome Variables*

<table>
<thead>
<tr>
<th>Measure</th>
<th>All Participants(^a)</th>
<th>Low(^b)</th>
<th>High(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
<td>(M)</td>
</tr>
<tr>
<td>L2 Fluency</td>
<td>139.44</td>
<td>19.73</td>
<td>122.23</td>
</tr>
<tr>
<td>Informal English</td>
<td>5.67</td>
<td>1.51</td>
<td>5.64</td>
</tr>
<tr>
<td>L1-L2 Match</td>
<td>6.51</td>
<td>3.87</td>
<td>7.27</td>
</tr>
<tr>
<td>L2 Commonality</td>
<td>0.16</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td>Consistency Index</td>
<td>1.13</td>
<td>0.32</td>
<td>1.17</td>
</tr>
</tbody>
</table>

\(^a\) \(n=45\); \(^b\) \(n=22\); \(^c\) \(n=23\)

As Table 3 demonstrates, participants on average scored just below 140 on L2 fluency and in the low to middle range on informal English (possible range: 4 - 12; observed range: 4 – 9). They produced about six L1-L2 lexical/conceptual matches out of 35 possible; followed native English speakers’ patterns with a frequency of .16; and produced more common associations when performing in Russian than in English as demonstrated by the mean of greater than 1.

At the group level, low L2 fluency speakers scored about 122 on L2 fluency and well above 5 on informal English. For the high L2 fluency group, the mean of L2 fluency was

\(^{10}\) All values have been rounded to two decimal places. When reporting exact \(p\)-values, those values have been rounded to three decimal places.
predictably higher, 155.91, but they scored about the same on informal English, 5.70. In other words, both groups showed low-to-mid range scores on informal English (possible range: 4 – 12). In terms of outcomes, low fluency speakers provided about seven L1-L2 conceptual matches; among the high L2 fluency group, the number was lower, about five. With regard to L2 commonality, low L2 fluency speakers showed a somewhat higher frequency compared to high fluency speakers, .18 and .14, respectively. Finally, on consistency measures, low L2 fluency participants scored higher than the high L2 fluency group, 1.17 vis-à-vis 1.08, indicating that among low L2 fluency speakers, there was less consistency across language conditions compared to high L2 fluency speakers.

Demographic (age of arrival), L2 proficiency (L2 fluency), and language experience (informal English) factors were examined for possible correlations with all three outcomes, see Table 4 below.

Table 4

<table>
<thead>
<tr>
<th>Predictor</th>
<th>L2 Fluency</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low(^a)</td>
<td>High(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AoA</td>
<td>-.21</td>
<td>-.29</td>
<td>.10</td>
<td>-.30</td>
<td>-.27</td>
</tr>
<tr>
<td>L2 Fluency</td>
<td>-.06</td>
<td>.38</td>
<td>-.18</td>
<td>.08</td>
<td>.16</td>
</tr>
<tr>
<td>Informal English</td>
<td>-.11</td>
<td>-.04</td>
<td>-.44(^*)</td>
<td>.05</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note. AoA = age of arrival; L2 Comm. = L2 Commonality; Cons. Ind. = Consistency Index.
\(^a\) n=22; \(^b\) n=23.
\(^*\) p < .05.

As shown in Table 4, age of arrival (AoA) did not result in any significant correlations with any of the three predictors in either group. (During preliminary analyses, it was found that
the two fluency groups were significantly different on AoA, a popular predictor in the literature; as a result, the AoA variable was included in the present set of correlational analyses.) There were no significant correlations of L2 fluency with any of the three outcomes in either group.

Finally, on informal English, where the two groups scored about the same (low L2 fluency: $M = 5.64$ ($SD = 1.43$); high L2 fluency: $M = 5.70$ ($SD = 1.61$)), there was a statistically significant negative relationship between consistency index and informal English among low L2 fluency bilinguals, $r(22) = -.44$, 95% CI [0.02, 0.73], $p < .05$. This result suggests that, with less use of informal English, low L2 fluency speakers tend to demonstrate a greater degree of inconsistency of associations across language conditions\textsuperscript{11}. All other correlations with this predictor did not reach statistical significance at the $p < .05$ level.

\textbf{Testing Inter-language and Intra-speaker Theories}

Two independent-samples $t$-tests were conducted. The two L2 fluency groups were compared on the number of L1-L2 matching pairs and L2 commonality scores to test for inter-language theory and intra-speaker theory, respectively. The results are presented in Table 5 below.
Table 5

Results of t-tests Comparing Low and High L2 Fluency Bilinguals on L1-L2 Match and L2 Commonality

<table>
<thead>
<tr>
<th>Outcome</th>
<th>L2 Fluency</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>t(43)</td>
<td>p</td>
<td>95% CI</td>
</tr>
<tr>
<td>L1-L2 Match</td>
<td>7.27</td>
<td>3.76</td>
<td>5.58</td>
<td>3.91</td>
<td>1.30</td>
<td>ns</td>
<td>-0.82, 3.80</td>
</tr>
<tr>
<td>L2 Commonality</td>
<td>.18</td>
<td>.06</td>
<td>.14</td>
<td>.06</td>
<td>1.93</td>
<td>ns</td>
<td>-0.00, 0.07</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval.

As Table 5 shows, there was not a statistically significant difference in the number of L1-L2 conceptual/lexical matches between low and high L2 fluency bilinguals, t(43) = 1.30, ns. In other words, high L2 fluency speakers did not differ from low L2 fluency speakers in their ability to access the conceptual storage. Similarly, no statistically significant difference was found between low and high fluency bilinguals with regard to their L2 commonality scores or how closely they follow native norms in their L2 associations, t(43) = 1.93, ns. Thus, in both instances, there was no evidence to support the alternative hypothesis (i.e., in both cases the corresponding confidence intervals include 0) so the null hypothesis could not be rejected.

General Cognitive Mechanism Hypothesis

Preliminary Analyses. A preliminary analysis of assumptions was performed to explore data normality and check for outliers. The outcome variable, consistency index, was examined for independence, linearity, homoscedasticity, and normality (Glass & Hopkins, 1996). All four assumptions were met12.

12 There was independence of residuals, as assessed by a Durbin-Watson statistic of 2.4. Next, a residual plot of the standardized residuals and standardized predicted values was produced; the residuals were scattered in a random manner, which suggested that linearity was present. Homoscedasticity was assessed based on a plot of the studentized residuals against the predicted values: the random shape of the plot indicated that the variable was homoscedastic. To test for normality, a P-P plot of the standardized residuals was produced; the residuals formed a diagonal line, demonstrating that the normality assumption was met.
Next, the outcome variable was inspected for outliers: there was one extreme value, 3.3 standard deviations greater than the mean. Outliers are common in research and present a particular problem especially when working with small samples\(^ {13} \). As Osborn & Overbay (2004) have pointed out, “1% of your subjects should be 3 standard deviations from the mean,” providing that an outlier is not data-entry error. Using a data transformation is one way to deal with outliers. I performed a square root transformation and then ran two regressions, with and without the transformation. When compared, the two models showed the same trends, which led me to keep the outlier. My decision was also based on the recognition that the ultimate goal of any statistical analysis is explanation. Any data transformation, on the other hand, makes subsequent explanation and interpretation of the results difficult (Tabachnick & Fidell, 2007).

**Statistical model-building.** My statistical model-building was guided by theory and previous research, both of which suggest that at different levels of L2 fluency bilinguals perform differently. In addition, studies have shown that quality of L2 experience, such as using English with friends who are native English speakers, impacts bilingual performance as well. Finally, the current study’s preliminary findings seem to indicate a possible interaction effect between L2 fluency and informal English: there was a significant negative correlation between informal English and consistency index among low fluency bilinguals \((r(22) = -.43, p < .05)\) but not high fluency bilinguals \((r(23) = .09, ns)\).

To investigate how L2 fluency, informal English, and their interaction affect bilingual performance, a three-step hierarchical model was tested. At steps one and two, L2 fluency and informal English were entered sequentially (Model 1 and Model 2, respectively); their interaction term was then entered at step three (Model 3), see Table 6.

\(^ {13} \) According to Cousineau and Charlier (2010), researchers working with smaller samples (e.g., \( N < 100 \)) should consider applying a more conservative approach, one that “[shows] a bias toward keeping the data” (p.60).
### Table 6

**Summary of Hierarchical Regression Analysis for Variables Predicting Consistency among Late-onset Bilinguals (N = 45)**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model 1 $\beta$</th>
<th>Model 2 $\beta$</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>T-Model $\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 Fluency</td>
<td>-.16</td>
<td>-.14</td>
<td>-0.00</td>
<td>0.00</td>
<td>-.12</td>
<td>-.10</td>
</tr>
<tr>
<td>Informal English</td>
<td>-.19</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.05</td>
<td>-.24</td>
<td>-.24</td>
</tr>
<tr>
<td>L2 Fluency x Informal English</td>
<td>0.00</td>
<td>0.00</td>
<td>.33*</td>
<td>.33*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Model Summary

<table>
<thead>
<tr>
<th></th>
<th>Model 1 $F$</th>
<th>Model 2 $F$</th>
<th>$R^2$</th>
<th>Cohen’s $f^2$</th>
<th>Change in $R^2$</th>
<th>$F$ for change in $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall $F$</td>
<td>1.15</td>
<td>1.36</td>
<td></td>
<td></td>
<td></td>
<td>2.76</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td>.03</td>
<td>.06</td>
<td></td>
<td>.17</td>
</tr>
<tr>
<td>Cohen’s $f^2$</td>
<td></td>
<td></td>
<td>.03</td>
<td>.06</td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td>Change in $R^2$</td>
<td></td>
<td></td>
<td>.03</td>
<td>.04</td>
<td></td>
<td>.11</td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
<td></td>
<td></td>
<td>1.15</td>
<td>1.55</td>
<td></td>
<td>5.28*</td>
</tr>
</tbody>
</table>

**Note:** T-Model = Model with transformed data. L2 fluency and Informal English were centered at their means.

*a – $f^2$: effect sizes calculated based on Soper (2014).

* $p < .05$.

As presented in Table 6, the two main effect models, Model 1 and Model 2, were not statistically significant, $F(1,43) = 1.15$, $ns$, and $F(2,42) = 1.36$, $ns$, respectively. Next, the interaction term was introduced (Model 3). By introducing the interaction term, a statistically significant increment of the amount of variance was explained ($R^2 = .11$, $F(1,41) = 5.28$, $p = .027$). As a result, the overall model showed a better fit with an increased value of $R^2$ of .17, which indicates a medium effect size, $f^2 = .20$ (Cohen, 1992). Overall, Model 3 was able to account for 17% of the variance in the measure of consistency. The direction of the beta weight ($\beta = .33$, $p = .027$), which had the opposite sign vis-à-vis the two main effects, indicated that the L2 fluency by informal English relationship became stronger as the levels of L2 fluency and informal English decreased.
In order to explore the interaction effect further, simple slope analysis was performed (Aiken & West, 1991). Regression lines were plotted and examined for two groups of informal English, low and high (i.e., 1 SD below and 1 SD above the mean), see Figure 1.

![Figure 1](image)

**Figure 1**
*Interaction of L2 Fluency and Informal English to Predict Consistency among Late-onset Bilinguals*

As Figure 1 demonstrates, the interaction between L2 fluency and informal English was crossover or disordinal (Aiken & West, 1991, pp. 22f). In this kind of interaction, the two predictors are not independent (Glass & Hopkins, 1996). In this particular case, the effect of L2 fluency on consistency was influenced or moderated by informal English. Specifically, in the interaction of L2 fluency and informal English predicting consistency, the regression slope for the low informal English group was significantly different from zero ($t = -2.28, p = .027$). In other words, L2 fluency was a statistically significant predictor of consistency under low
informal English: higher L2 fluency scores associated with lower values of consistency index (i.e., trending toward 1), indicating less discrepancy between L1 and L2 associations. However, under high informal English, L2 fluency was not a significant predictor: the regression slope for the high informal English group was not significantly different from 0 ($t = 0.98$, ns). That is, there was no evidence of L2 fluency’s affecting consistency among high informal English bilinguals.

**Discussion**

The primary goal of this study was to test several interpretations of linguistic relativity based on a sample of late-onset Russian-English bilinguals and examine the role of L2 experience within a context of LR debates. Overall, it was found that language experience (operationalized as use of informal English) plays an important role in bilingual performance. A statistically significant interaction between L2 fluency and informal English was found indicating that language experience moderates the effects of L2 fluency. Among bilinguals who did not use English in informal settings often, L2 fluency predicted whether their L2 associations were similar to and consistent with their L1 associations. In contrast, among those bilinguals who used informal English often, L2 fluency was not a significant predictor of consistency.

The study utilized a popular word-association paradigm to test several interpretations of LR. Each interpretative framework generated its own prediction about influences of language on bilingual performance and thus provided opportunity to gather empirical evidence for or against each theory within a context of one task. Notably, each theoretical framework ascribes importance to different factors, e.g., L1 or language experience. Viewing language as representational and a-temporal in nature, inter-language theory posits that L1 exerts its influence across language conditions. Intra-speaker theory, on the other hand, goes a step further:
it does not dispute that language is internally represented, but brings the speaker into the picture to explain how change happens and why L1 may not have privileged status. In contrast, general cognitive mechanism theory does not make any claims about the status of language within a speaker; rather it views language in terms of cognitive processes and language experience.

There was no evidence in support of inter-language theory: less and more proficient bilinguals did not differ on the number of L1-L2 conceptual matches. Contrary to Kroll and Stewart’s (1994) model, more fluent bilinguals did not seem to have better access to the conceptual storage as the difference in the number of L1-L2 matching pairs among less and more fluent bilinguals was not statistically significant ($t(43) = 1.30, ns$). One potential limitation in this regard, though, would be a possible priming effect for English. In other words, applying Grosjean’s (2001) notion of language mode, one might ask: Could it be that the bilinguals were in a predominately L2 mode and thus completed the Russian (L1) part under the influence of English (L2), which then resulted in the negative finding for the inter-language hypothesis? The issue of controlling for either L1 or L2 has been examined in the literature extensively: many researchers agree that even when a researcher sets up her study so that all instructions are carried out in only one language and bilingualism is never mentioned, it is virtually impossible to ensure a completely monolingual (L1) mode, especially when the study takes place in a host (L2) country (Mirian & Spivey, 2003). Furthermore, Grosjean’s theory of language mode asserts that there is an L1-L2 continuum that underlies a bilingual’s performance, which seems to suggest that a completely monolingual mode (either L1 or L2) is an abstraction. Rather, we should expect that by inducing a bilingual mode with L1 as the base language we could control for any L2 priming effect and this is what was done in the present study: all participants were greeted and oriented to the study in Russian. Next, they read task instructions in Russian, completed a 5-
word trial series in Russian, and then proceeded to do the experimental list in Russian. All in all, every attempt was made to “nudge” the participants away from the L2 end of the continuum and activate L1 so that they would be in a bilingual Russian-based mode during their performance of the Russian version of the task. Given that the participants were in a Russian-based bilingual mode, that is, had both languages activated with Russian serving as their base language, it is unlikely that there was a priming effect for English and, thus, the negative finding should stand.

Next, in regards to intra-speaker theory, no evidence was found to corroborate it: there was no difference in native-like (L2) commonality of association between less- and more-fluent bilinguals ($t(43) = 1.93, n.s.$). This finding contradicts the studies by Zareva and colleagues (Zareva, 2010; Zareva et al., 2005) which found that more proficient bilinguals tended to resemble native speakers more compared to less proficient bilinguals. The difference in findings may have to do with differences in stimulus selection and L2 proficiency measures. In the present study, the list was based on common, everyday words and L2 proficiency was measured via an on-line test of spoken English. Zareva and colleagues, however, used words of various frequencies, excluding high-frequency words and including low-frequency words, which are normally not part of everyday discourse. In addition, the researchers assessed L2 proficiency based on a written test of L2 vocabulary. As a result, the L2 proficiency data, as well as the final findings, may have been reflective of different levels of formal English instruction or educational attainment, which would make the L2 proficiency factor rather circumscribed. Finally, the present finding confirmed Fitzpatrick’s (2007) reservation about the use of native norms as a means to gauge bilingual development: there was not a trend toward more native-like associations among more proficient bilinguals.
One caveat is in order here: it concerns a potential effect of word selection on the intra-speaker findings. In particular, the list was based on well-known high frequency words which might have produced a ceiling effect, which in turn might have resulted in the finding of no support for intra-speaker theory. My original word selection process was guided by an important design consideration: I had to ensure that all cue words would have the corresponding native L2 norms; the L2 dataset used in the study was based on high frequency words. Future studies should use words of different frequency to minimize the risk of ceiling effect.

While neither of the two representational frameworks received support, there was support for GCM theory in that language experience was found to play a role in bilingual performance: overall, as L2 fluency and informal English tended to decrease, the effect of their interaction on consistency tended to increase. In particular, among infrequent users of informal English, consistency was predicted by L2 fluency while among bilinguals, who were frequent users of informal English, there was no relationship between L2 fluency and consistency of associations. This finding adds to the existing body of research that has found that language use is important in “determining the L2 acquisition end state” (Birdsong & Molis, 2001, p. 247; see also Moyer, 2011).

The present study also provides support to the claim advanced by Fitzpatrick (2007) that speakers (whether mono- or bilingual) should follow a preferred association style. As I found, among more fluent bilinguals and those with higher levels of informal English, consistency scores tended to hover around 1, indicating that these bilinguals consistently produced associations similar (or dissimilar) to that of their peers in both languages. This finding also supports the view that bilinguals are not “two persons in one body” (Grosjean, 1989); thus, we
should consider using similar models whether we study language acquisition in monolingual speakers or bilingual development (e.g., Brysbaert & Duyck, 2010).  

As far as potential limitations are concerned, gender differences could have affected the study’s findings: there were three times as many female as male participants. Indeed, there seem to be gender differences in L2 acquisition, although the findings have been inconclusive (Huerta Tong, Irby, & Garza, 2012). So far, gender differences in SLA have been modeled after research on first language acquisition showing an advantage among girls that starts at an early age and seems to persist throughout school years (Burman, Bitan, & Booth, 2008; Roulstone, Loader, Northstone, & Beveridge, 2002). It has not been clear if gender plays a role due to its biological nature or social and cultural factors. Future studies should consider adding gender among their variables of interest.

Another potential limitation of the study was its small sample size, which might have impacted some of the analyses and made the statistical tests less powerful. For example, none of the main effects in the regression model were statistically significant. Also, having to deal with an outlier was likely another consequence of the small sample size. It must be noted though, that given the exploratory nature of this study, it was difficult to guesstimate the desirable effect size or what kind of statistical analysis would be appropriate, which are just some of the factors that determine one’s choice of power (in addition to the traditional concerns about sample size) (Cohen, 1992; Wampold & Freund, 1987). The task was not easy because most researchers who use word-association tasks with bilinguals rely on correlations, t-tests, and one-way ANOVA’s as their main statistical tools and very rarely report effect sizes. Their samples range from 17 participants per condition (e.g., Zareva et al., 2005) to 30 participants per study (Fitzpatrick, 2007). In this study, 45 participants were grouped into two fluency levels, low and high, with 22
and 23 participants in each group, which was comparable to other studies within this tradition. Also, using a more powerful statistical tool, regression, allowed for the testing of rather complex predictions even with this modest sample size.

One more potential limitation had to do with one of the predictors having a non-normal distribution where a sizable group of scores clustered at the higher end of the distribution. The predictor, L2 fluency, was based on two subscales from the Versant Test of Spoken English, a standardized test of English with strong reliability and validity (Bernstein et al., 2009). However, in both subscales, the range was not sensitive enough to accommodate a sample of bilingual speakers with a high proportion of very advanced bilinguals. In the sample of 45 bilinguals, 13 participants scored at a ceiling on both fluency and pronunciation and as a result had the maximum score of 160. This non-normality might have influenced the regression model and could have been another reason why none of the main effects in the model turned out to be significant. Alternatively, the absence of statistically significant main effects could be due to the nature of the underlying ANOVA model (Aiken & West, 1991): models with a significant interaction term which is disordinal, may have main effects that are not significant, yet the interaction term may be large and significant (Lubin, 1961). Future studies would have to address this issue of significant disordinal interaction through experiment in order to answer the following two important questions: “How does this interaction occur? How can I bring it under experimental control?” (Lubin, 1961, p. 816).

The present study has other implications for future research in that it shows the importance of including measures of language experience when conducting research with late-onset bilingual adults. Future studies should try to isolate what makes language experience an important factor in bilingual performance: whether it is the very act of social engagement that
benefits bilinguals; or the variety of L2 input in terms of L2 vocabulary, syntax, or different registers; or the actual process of speaking and interacting; or some combination of these factors. We now have a growing body of research that highlights the importance of social and cultural context in human development: instead of looking for prototypical, universal laws of human behavior along either linguistic or cognitive lines, this tradition pays attention to how people use language to accomplish their everyday tasks and find themselves within their cultural time and place (Brockmeier, 2012). After all, what’s the point of acquiring a language, first or second or third, if not to join a community of minds (Nelson, 2007)?

These non-representational, experiential approaches point out that language is more than language processing because speakers use language in very non-linear and fragmented ways. (For example, speakers tend to use one-word responses when mutual understanding is assumed; they turn to extended clausal responses when communication breaks down (Fox & Thompson, 2010).) That is, as speakers tell their funny, sad, profound, or very trivial stories, they anticipate how others might react to their story and may need to “re-write” their story on the fly; the others, in turn, may interject, or take over the story and then let the speaker finish or not. In other words, the point of language is not so much in conveying information but in the experience of active co-participation (Brockmeier, 2012). According to this view, exploring how bilinguals become active members of their new community of minds and taking a close look at their language experience can shed new light on the very nature of language. After all, words, phrases and phonemes are linguistic projections (Brockmeier & Harré, 2001) and we need to be as close to the actual experience of a speaker as we can.

Also, given the ever-growing interest in bilingual studies across the lifespan (Bialystok, 2005), if we include language experience in bilingual research, that could potentially allow for
new insights on the sources of bilingual advantage among bilinguals compared to monolingual adult speakers. Once we have isolated and clarified the mechanism through which language experience impacts bilinguals, we could develop interventions for bilingual as well as monolingual speakers who thus might also benefit from the research on bilingual advantage.

With regard to implications for instruction, the present study provides support to the growing recognition among second language instructors that fluency and language use are important factors in second-language acquisition (e.g., Gilbert, 2008). Although formal instruction in English literacy is crucial to second-language learners’ success in today’s workplace, informal language use can provide additional benefits as this study demonstrates: among infrequent users of informal English, the relation between L2 fluency and informal English was predictive of bilingual performance.

In closing, this study challenges representational approaches to language, which tend to regard language as an innate capacity, overemphasize the importance of L1 and thus view bilingual development as built around L1; as importantly, they minimize social and cultural factors. In contrast, the present study demonstrates that a stronger focus on language experience can enhance our understanding of language and cognition and enable us to make predictions about performance among bilingual adults based on L2 proficiency and language use. In a word, paying more attention to language experience holds the promise of making Whorf’s view on language, and our research within this tradition, more relevant to the needs of bilingual speakers and the community at large.
CHAPTER TWO: STUDY TWO

FROM THOUGHT AND LANGUAGE TO LANGUAGE IN CONTEXT: REEXAMINING LINGUISTIC RELATIVITY IN A SAMPLE OF LATE-ONSET BILINGUALS
Introduction

Whorf’s claim that differences between various groups of people can be explained in terms of the language they speak (linguistic relativity or LR) has been recently revived across many fields, including cognitive psychology, psycholinguistics, and research on second-language acquisition. The resulting scholarship has produced diverse findings and readings of LR. For example, Reines and Prinz (2009) have proposed that one way to interpret LR would be Habitual Whorfianism, or the view that different languages instill in their speakers habitual ways of speaking and thus thinking. As some proponents of Habitual Whorfianism would say, language, through its grammar, provides a lens that shapes and colors our thinking (see Slobin’s (1996) notion of speaking for thinking). For example, Boroditsky and Schmidt (2000) have examined LR with groups of bilingual speakers. Comparing speakers of languages that have a grammatical gender category, they found that native German and Spanish speakers tend to show a gender bias when they use a non-gendered language such as English. In particular, German and Spanish speakers were asked to describe inanimate objects by producing three English adjectives (all objects were of opposite gender in German and Spanish). A group of native monolingual English speakers rated the adjectives as masculine or feminine. Spanish speakers were found to use more masculine adjectives in describing objects that are of masculine gender in Spanish and more feminine for objects that are feminine in Spanish. The same pattern was found for the German group. The researchers concluded that the observed masculinity bias thus demonstrated that a speaker’s L1 grammar influences their cognitive performance in an L2 context.

In a related line of research, monolingual adult and child speakers of non-gendered (English) and gendered languages were shown pictures of inanimate objects and asked to circle either M or F based on whether they thought the picture looked “masculine” or “feminine” (Sera,
Berge, & Pintado, 1994); next, adult speakers of Spanish, French, and German were asked to assign male or female voices to pictures of humans and inanimate (artificial and natural) objects (Sera et al., 2002). In both studies, the researchers found some evidence of grammatical gender affecting classification behavior but the effect was not uniform. For example, there was a main effect of language among speakers of Spanish and French, who tended to assign voices based on the grammatical gender of their language; however, German speakers did not classify stimuli according to their gender system (Sera et al., 2002). Sera and colleagues surmised that French and Spanish have many formal markings for grammatical gender; in addition, in Spanish, grammatical gender seems to correspond with natural gender. In German, on the other hand, there is less correspondence between grammatical and natural gender and fewer formal markings, all of which may contribute to a lack of correspondence between grammatical gender and classification among this group of speakers. The authors concluded that “the relation between grammatical gender and classification suggested by the literature varies as a function of language and the task that is used to examine the relation” (p. 379)\(^\text{14}\). (In addition, the relation between grammatical gender and classification would be expected to vary as a function of how grammatical gender is organized in any particular language: some languages base their gender systems on sex while others on animate/inanimate or human/non-human.) Overall, Sera and colleagues’ research exemplifies linguistic relativity as an inter-language phenomenon, i.e., speakers of different languages show different patterns of performance in L2 contexts due to differences in their respective L1 grammars.

\(^{14}\) Sera et al. (1994) review three different readings of gender: natural (universal distinction between male and female); conceptual (e.g., artifacts – male, natural objects – female, which may also be universal, according to some proposals; linguistic: semantic (in English: gender part of meaning, e.g., *hen*) and grammatical (French, Russian: all nouns are gendered; gender expressed formally, e.g., suffix, inflections, etc.).
More recent studies have questioned the received view that a bilingual’s L1 exerts stable, predictable influence on her performance in L2 contexts. Challenging early “gender-bias” research, Kousta, Vinson, and Vigliocco (2008) conducted a study with English and Italian monolingual speakers and a group of highly proficient early-onset Italian-English bilinguals. The researchers tested the claim that grammatical gender influences conceptual associations in Italian-English bilinguals when they are tested in English (Italian has a grammatical gender system while English does not). All participants were given an error induction test based on a series of 26 pictures of common animals. Stimuli (drawings) were presented on a computer screen, at a fast rate, and participants were to name the presented animal. Errors were then collected and analyzed for gender bias. It was found that Italian-English bilinguals tend to make errors that preserve the gender of a noun when tested in Italian (which was similar to how native Italian speakers performed) but not when tested in English. The researchers concluded that “gender bias effects are very constrained, at best”: there is no evidence of LR as *inter-language relativity* but rather there is evidence of *intra-speaker relativity*, i.e., bilingual speakers follow either L1 or L2 patterns, depending on task context, which ultimately suggests that grammatical gender does not have a conceptual effect.

Also, Athanasopoulos and Kasai (2008) looked at how languages with different plural markings may affect non-linguistic categorization in bilinguals. They compared more and less proficient Japanese-English speakers (as based on a test of English vocabulary) to monolingual Japanese and English speakers in a classification task. All participants were asked to classify triads of novel objects as similar in either color or shape. The findings seem to support the intra-

---

15 English proficiency was based on self-reports. Most bilinguals were early-onset bilinguals (*M* _age of acquisition_ = 10.1, _SD_ = 2.9) with English being the predominant language of use (*M* _% language use per day_: English, 70 _%_ (_SD_ = 20%); Italian, 30 _%_, _SD_ = 20%).

16 Task context means language of task instruction, i.e., English or Italian.
speaker view of LR: it was found that more proficient bilinguals showed a pattern similar to that of English monolinguals’ (shape preference), regardless of whether they were tested in English or Japanese. The less proficient group was similar to Japanese monolinguals (color preference).

The two versions of LR, inter-language and intra-speaker, share their view of language as representational in nature, whereby language’s primary function is to reflect and accurately represent the world, primarily through the use of grammar. Though intra-speaker relativity allows for development and change within individual speakers, its main focus is still on recording change rather than looking for the mechanisms of change. On the other hand, non-representational accounts of language are concerned with the ways language is used, its social and communicative functions (Tomasello, 2003), and how different linguistic and non-linguistic factors work in concert to produce the dynamic phenomenon that is language. From a functionalist’s perspective, grammatical categories are pairings of form and meaning as practiced by communities of speakers, patterns that come into being, become visible and predictable and then dissolve due to interplay of internal and external factors (e.g., Ford, Fox, Thompson, 2002).

Accordingly to Beckner et al.’s (2009) view, language is best viewed as “the cognitive organization of language experience” [emphasis added] (see also Bybee, 2006). Broadly speaking, language experience refers to the totality of local speech events that occur in particular historical and cultural settings. Next, the term “organization” refers not to innate structures of universal grammar but to general cognitive mechanisms that shape and organize our dynamic and variable experience, such as sequential processing and pattern recognition (Beckner et al., 2009).

In a similar vein, cultural psychology views language as a means of accomplishing meaningful social and cultural goals at both group and individual levels (Vygotsky, 1978). When
we acquire language, we gain entry into a “community of minds” (Nelson, 2007). Also, given its non-linear and fragmented nature as it is practiced in everyday discourse (Brockmeier, 2012), language has to be studied along with other, non-linguistic factors (e.g., social, cultural, historical). If we limit ourselves to the view of language as an exclusively linguistic phenomenon and then hope to explain human behavior by mapping linguistic concepts onto cognitive concepts, we will be committing an “ontological fallacy” (Brockmeier & Harré, 2001).

Applied researchers who study second language acquisition have, too, long recognized the importance of language experience (e.g., L1 or L2 use in different settings with different groups of speakers). There have been studies about L1 and L2 use among bilinguals and how language experience impacts L2 proficiency (often operationalized as L2 accent). Flege and colleagues have examined biological (age) and experience-based factors (e.g., length of residence, frequency of L1 and L2 use) and found language experience to be a predictor of L2 proficiency: among early-onset bilinguals (age of onset < 16), a bilingual’s L1 use was found to be negatively correlated with L2 proficiency (measured as English accent) (Flege, Frieda, & Nozawa, 1997); among late-onset bilinguals (age of onset > 16), length of residence was positively correlated with L2 proficiency only if an L2 was used on a regular basis (Flege & Liu, 2001). It is important to mention that second language researchers prefer nonverbal measures of L2 proficiency like L2 accent to the traditional linguistic measures of L2 vocabulary and L2 syntax when they study the effects of biological or experiential factors (e.g., age or language experience) on L2 competence. This preference of nonverbal over verbal measures seems to suggest a certain degree of tension between applied and theoretically-oriented researchers and a need for a more integrated and comprehensive approach to the study of bilingual development, and language and cognition in general.
Study Description

The present study set out to test three interpretations of LR in a bilingual context to explore the way language and cognition relate to each other. Similar to Boroditsky and Schmidt’s (2000) study, a word-association task was given to a group of late-onset Russian-English bilinguals at two levels of L2 proficiency; their language experience with L1 and L2 was assessed via survey. Participants were to generate adjectives for 3 clusters of nouns (five nouns per cluster), corresponding to the three grammatical genders in the Russian language (masculine, feminine, and neuter). Their responses were rated for masculinity by a group of native English speakers and then examined for evidence of L1’s influences under L2 conditions (masculinity bias) according to the following three frameworks:

**Inter-Language Theory.** According to this interpretation, when bilingual speakers switch between languages, their first language continues to influence their cognitive performance under L2 conditions, purportedly because language instills in its speakers certain *habitual* ways of thinking, according to Habitual Whorfianism (Reines & Prinz, 2011). For example, L1 grammatical categories such as gender may affect a bilingual’s judgments of masculinity in an L2 (e.g., Boroditsky & Schmidt, 2000). In the present study, inter-language theory predicted that associations for masculine clusters should receive higher masculinity scores than either feminine or neuter clusters and thus produce evidence of masculinity bias among Russian-English bilinguals.

**Intra-Speaker Theory.** Although built upon Habitual Whorfianism, this reading of LR suggests a more flexible link between L1 and L2: it proposes that, at higher levels of L2 proficiency, bilinguals tend to follow native patterns of either L1 or L2 monolinguals (e.g., Kousta et al., 2008). In the present study, this framework predicted that there would be
differences in masculinity scores among clusters between more and less proficient bilinguals so that less proficient bilinguals would show evidence of masculinity bias. Also, there would be no evidence of masculinity bias among more proficient bilinguals as their associations should receive similar masculinity scores across clusters (as one would expect to find no evidence of masculinity bias in a native English speaker’s performance).

**General Cognitive Mechanism (GCM) Theory.** This non-representational interpretation of LR predicted the opposite of the other two theories: since language is not reducible to a-temporal, innate structures and linguistic and cognitive concepts do not necessarily match (e.g., grammatical gender vis-à-vis natural gender), responses should not differ as a function of noun cluster or L2 proficiency. That is, there should be no difference in masculinity scores between masculine, feminine, and neuter clusters in both groups of bilinguals, more and less proficient. Regarding native speakers’ scoring, the GCM framework predicted significant agreement among raters as a result of shared cultural practice.

**Method**

**Participants**

Forty-five late-onset Russian-English bilingual speakers participated in the study\(^\text{17}\). To be eligible for the research, they had to meet the following criteria: they were expected to be functionally fluent in and speak both languages in a variety of settings (e.g., home, work, with family and friends, etc.). Furthermore, they were expected to have spoken Russian as their first language and have not had significant exposure to ambient English prior to coming to the US at age 20 or later, after graduating high school in Russia (or their native country, a former Soviet republic). Finally, they were expected not to have had any history of learning disability. All

\(^\text{17}\) This study was run with the same group of bilinguals as study 1.
participants went through a screening phone interview in Russian to assess their command of spoken Russian.

Participants were recruited via fliers, internet postings, and word of mouth. For their participation, they received small gift cards ($20) and a copy of their English proficiency test, which was part of the study and included strategies to improve English.

**Procedure and Instrumentation**

All participants were tested individually in a psychology lab by the author. First, participants were greeted and instructed about the study in Russian in order to induce a bilingual mode (Grosjean, 2001). They then completed several tasks and language measures in both languages: a one-word association task in Russian, a language experience survey in English, an English version of the one-word association task, an L1-gender association task in English and, finally, a standardized test of spoken English (the data of the one-word association tasks in Russian and English were not part of the current study). All measures were untimed. It took participants, on average, about one hour to run the complete protocol.

**L1-Gender Association Task.** The L1-gender association task consisted of a series of 15 stimulus words (see Cue Words below) presented on a computer screen, one cue at a time, in random order. The task was administered in English; it was not timed, instead participants were asked to work at a brisk pace. The task was based on a list of 15 cue words – common English nouns describing inanimate nature-related phenomena such as “flower” and “thunder” (see Appendix E). There were an equal number of nouns that are ascribed masculine, feminine, or neuter grammatical gender in Russian (i.e., each gender cluster thus comprised five cues).

---

18 According to Grosjean (2001, 2008), bilinguals use L1 and L2 on a monolingual-bilingual continuum whereby they may rely more on either L1 or L2 as their base language and thus activate the base language (L1 or L2) to such a degree that would qualify them as being in a monolingual mode. Introducing L1 in an otherwise L2-predominant environment is said to induce a bilingual mode.

19 The one-word association task (in Russian and English) and its data were analyzed in study 1.
Participants were instructed to read each cue out aloud, say an adjective associated with the cue, and then proceed to the next screen by mouse click. After completing a series of three trial cues, they completed the 15-word series. Their audio responses were recorded and transcribed. The generated adjectives were then pooled together and checked against the *Oxford English-Russian Bilingual Dictionary* (Oxford Language Dictionaries Online, n.d.) to verify their part of speech status: sixty two adjectives were judged to be proper English adjectives as indicated by the dictionary. A group of seven native English speakers rated the 62 adjectives for masculinity (see Appendix F). Each participant then received three masculinity scores, one for each gender cluster (i.e., masculine, feminine, and neuter), calculated as the sum of masculinity points awarded for each proper adjective in that cluster; possible range: 0 – 35, whereby 0 indicated no masculinity points for the cluster, and 35 indicated that all 5 adjectives in the cluster received the highest masculinity score of 7 (see Masculinity Scores).

**Cue Words.** Fifteen common nouns (Appendix E) were selected for their semantic and translation properties. The decision to choose inanimate nouns referencing nature-related phenomena was made in an attempt to expand previous studies. In particular, some researchers have found that speakers of “gendered” languages tend to associate English nouns referencing tools and man-made objects with adjectives rated by native English speakers as masculine (Boroditsky & Schmidt, 2000). Others have observed stable classification patterns among monolingual English speakers whereby artificial and natural objects would be classified along masculine and feminine lines respectively. Additionally, monolingual Spanish speakers tended to classify objects according to the grammatical gender of the Spanish language (Sera et al., 1994). To this author’s knowledge there have been no L1-gender association studies using nouns referencing inanimate, not man-made phenomena.
With regard to translational properties of the selected cues, each cue was checked for its translational equivalency as customary among researchers working with word association tasks (e.g., Prior, MacWhinney, & Kroll, 2007). All 15 testing words and three trial words had a single or dominant translation in Russian based on translations provided by four Russian-English bilinguals who were not part of the study. In addition, all stimulus words were checked against the Oxford English-Russian Bilingual Dictionary to confirm that they only had one or dominant translation in Russian.

Test of Spoken English. To assess for L2 proficiency, participants completed a standardized test of spoken English (http://www.versanttest.com): the Versant Test of Spoken English is administered over the phone and instantly produces several measures. Designed to assess speakers’ facility with spoken English, the test is commonly used by academic and applied researchers working with bilinguals (e.g., Bradlow et al., 2011; Douglas, 2013). The test provides one overall score and four sub-scores based on four subscales that cover two broad areas, content and control (subscale range: 20 – 80). The former refers to the ability of a speaker to understand spoken English and thus taps into their grasp of English grammar and vocabulary. The latter refers to the ability of a speaker to express themselves clearly in spoken English. There are two subscales to measure a speaker’s facility with English grammar and the lexicon: Sentence Mastery and Vocabulary, respectively. Sentence Mastery examines the ability to process and produce English phrases and clauses while Vocabulary reflects the ability to use common everyday words. The test has strong reliability and validity: the split-half reliability coefficients for the two sub-scales range from .92 to .95 (machine and human scored, respectively). Validation studies have shown that Versant test scores highly correlate with other measures of English proficiency such as the TOEFL (Bernstein, Van Moere, & Cheng, 2010).
Given that both subscales, Sentence Mastery and Vocabulary, assess related areas (content) and the two subscales highly correlate with each other \((r=.72)\), the two sub-scores were combined to produce one score as a measure of on-line spoken proficiency in English (see L2 Content under Measures below).

**Language Experience Survey.** The Language Experience Survey was adapted from the Li, Sepanski, & Zhao (2006) questionnaire (see Appendix D). The questionnaire was selected because it covers a broad range of demographics and experiential measures: the survey included sections on demographics, self-ratings of English and Russian proficiency across modalities, estimates of language use in different social settings and for different purposes, and others. Only the sections on demographics and language use in English and Russian informal settings were used in the current study (for a detailed description of language use data, see Measures, Informal English).

**Measures**

**Masculinity Scores.** After verifying all descriptors for their status as English adjectives (Oxford Language Dictionaries Online, n.d.), all descriptors that qualified as proper English adjectives were pooled together thus resulting in a list of 62 adjectives (Appendix G). A group of seven monolingual native English speakers then rated each adjective as either masculine (1) or feminine (0) (see Appendix F). Interrater reliability was estimated using an intraclass correlation coefficient (ICC): the ICC was .87 (95% CI [.82, .92]). Based on the native English speakers’ ratings, each adjective received a masculinity score calculated as the sum of all seven ratings for that adjective (range: 0 – 7). All participants then received three masculinity scores, one for each gender cluster – feminine, masculine, and neuter cluster scores – calculated as the sum of masculinity ratings for all qualified adjectives in this cluster (range: 0 – 35). Additionally, for
each participant, two masculinity bias scores were calculated as the difference between masculine and feminine cluster scores, and masculine and neuter cluster scores (Masculinity Bias 1 and Masculinity Bias 2 respectively; range: -35 – 35). For example, a score of 35 would indicate that all five adjectives from the masculine cluster received the highest score of 7 while all five adjectives from the feminine cluster received the lowest score of 0 and, thus, the bilingual’s performance demonstrates a perfect masculine bias, as measured by masculinity bias 1. A score closer to 0 would indicate very little difference between masculine and feminine clusters and suggest no evidence of masculinity bias; ditto for masculinity bias 2.

**L2 Content.** To measure L2 proficiency, an L2 Content score was calculated for each participant as the sum of their sentence mastery and vocabulary sub-scores on the Versant test of spoken English (possible range: 80 to 160, where 20 is the lowest and 80 the highest score on each subscale).

**Informal English.** Participants’ use of the English language was measured using four items from the language experience survey. The items asked participants to estimate how much time they spend speaking English with their spouses, children, parents and relatives, and friends. To calculate Informal English scores, a composite variable (Informal English) was created after subjecting four Informal English sub-scores/items to a series of data reduction procedures. First, to normalize the data, all responses were re-coded as infrequent (1), moderate (2) or frequent (3). Then a composite Informal English score for each participant was calculated as the sum of their four sub-scores in the Informal English category (i.e., (i) spouse or partner, (ii) children, (iii) other family members, and (iv) friends). This re-coding process produced a scale with a possible range of 4 to 12. The same kind of re-coding procedure was applied to calculate Informal Russian scores. The inclusion of this additional variable was necessary to explore whether
informal English correlates with informal Russian. A significantly high level of correlation would be problematic as it would suggest that a speaker’s interpersonal style might be a factor. To examine whether informal English was representative of language experience and not indicative of interpersonal style, the two variables, informal English and informal Russian, were examined at the overall group level. The correlation coefficient was not significant at the $p < .05$ level, indicating that informal English was a measure of language use and not a measure of interpersonal style.

Results

Descriptive Findings

To allow testing of inter-language theory and intra-speaker theory, two groups were created based on a median split ($M_{group} = 138.40$, $SD = 19.06$), which allowed to create two virtually equal groups: bilinguals with L2 content scores equal to or lower than 143 were classified as low L2 content bilinguals and those with scores equal to or greater than 145 as high L2 content. Next, a series of $t$-tests were performed to assess for any differences in demographic characteristics and patterns of language use between low and high L2 content bilinguals. The means and standard deviations of demographic and language use measures are presented in Table 7.
### Table 7

**Means (M), Standard Deviations (SD), and t-test Results for the Demographic and Language Use Variables by L2 Content**

<table>
<thead>
<tr>
<th>Variable</th>
<th>All participants&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Low&lt;sup&gt;b&lt;/sup&gt;</th>
<th>High&lt;sup&gt;c&lt;/sup&gt;</th>
<th>t(43)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Age</td>
<td>37.56</td>
<td>9.44</td>
<td>37.81</td>
<td>11.03</td>
</tr>
<tr>
<td>AoA</td>
<td>26.32</td>
<td>6.27</td>
<td>27.91</td>
<td>7.59</td>
</tr>
<tr>
<td>LoR</td>
<td>11.23</td>
<td>6.98</td>
<td>9.91</td>
<td>8.23</td>
</tr>
<tr>
<td>EDU</td>
<td>3.53</td>
<td>0.92</td>
<td>3.27</td>
<td>1.12</td>
</tr>
<tr>
<td>IEng</td>
<td>5.67</td>
<td>1.51</td>
<td>4.41</td>
<td>0.50</td>
</tr>
<tr>
<td>IRuss</td>
<td>7.20</td>
<td>2.37</td>
<td>7.45</td>
<td>2.37</td>
</tr>
<tr>
<td>Gender</td>
<td>35f, 10m</td>
<td></td>
<td>16f, 6m</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Age = chronological age, in years; AoA = age of arrival in the United States; LoR = length of residence in the United States; EDU = educational attainment. EDU scores range from 0 (Did not Graduate High School) to 5 (Completed Post-Graduate Degree). IEng = Informal English, scores range from 4 (infrequent use of English with spouse, children, friends and family) to 9 (moderate to frequent use of English with spouse, children, friends and family). IRuss = Informal Russian, scores range from 4 (infrequent use of Russian with spouse, children, friends and family) to 12 (frequent use of Russian with spouse, children, friends and family).

<sup>a</sup>n=45;  <sup>b</sup>n=22;  <sup>c</sup>n=23

* p < .01

As Table 7 shows, there were no significant differences between low and high L2 content participants with regard to their age, age of arrival, length of residence, and educational attainment. In their uses of the two languages in informal social settings, the two groups did not differ with regard to their uses of the Russian language but did differ in their uses of English: low L2 content bilinguals scored 2.46 points lower than the high L2 content group (p < .01) which means that the low L2 content group used English in informal settings significantly less frequently compared to the high L2 content group.

**Preliminary Findings**

On average, masculinity scores across clusters ranged from 11.87 to 21.40; masculinity bias measures ranged from -9.53 to -1.82 for all participants, see Table 8.
Table 8

*Means (M) and Standard Deviations (SD) of the Outcome Variables*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>All participants&lt;sup&gt;a&lt;/sup&gt;</th>
<th>L2 Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Masculinity by Cluster*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masculine</td>
<td>11.87</td>
<td>5.74</td>
</tr>
<tr>
<td>Neuter</td>
<td>13.69</td>
<td>5.47</td>
</tr>
<tr>
<td>Masculinity Bias**, (MB):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB, II</td>
<td>-1.82</td>
<td>7.76</td>
</tr>
</tbody>
</table>

*Note.* MB, I = Masculinity Bias, 1 (Masculine v. Feminine Cluster); MB, II = Masculinity Bias, II (Masculine v. Neuter Cluster).

*Possible range: 0 – 35; observed range: 0 – 30.

**Possible range: -35 – 35; observed range: -28 – 14.

<sup>a</sup>n=45; <sup>b</sup>n=22; <sup>c</sup>n=23

As demonstrated in Table 8, at the overall group level as well as across L2 content levels, feminine clusters had the highest masculinity scores of about 21; masculine and neuter clusters, on the other hand, hovered in the lower teens from about 11 to 15, at the highest. Next, across L2 fluency levels, masculinity bias I scores were higher (in absolute values) than masculinity bias II scores. Moreover, the masculinity bias I values had a negative sign, indicating that feminine clusters had higher masculinity scores than masculine clusters.

**Testing Inter-language Theory**

To examine for evidence of L1 grammar’s influencing bilingual performance under L2 conditions (i.e., masculinity bias), two one-sample *t*-tests on the masculinity bias scores were conducted (Masculinity Bias I and Masculinity Bias II). There was a statistically significant difference in masculinity scores between masculine and feminine clusters (Masculinity Bias I): feminine clusters, on average, received higher masculinity scores than masculine clusters: the mean of -9.53 (95% CI, -11.64 to -7.43) was significantly different from zero, *t*(44) = -9.14, *p* <
.01. The negative sign of Masculinity Bias I thus indicated a “reverse” masculinity bias (see Figure 1 below). There was no difference in masculinity scores between masculine and neuter clusters (Masculinity Bias, II): the mean of -1.58 (95% CI, -0.83, 0.10) was not significantly different from zero so there was no evidence of masculinity bias when masculine clusters were compared to neuter clusters, $t(44) = -1.58, p = .122$.

Figure 2

"Reverse" Masculinity Bias

Note. MB, I = Masculinity Bias, I (Masculine Cluster v. Feminine Cluster); MB, II = Masculinity Bias, II (Masculine Cluster v. Neuter Cluster).

As Figure 2 illustrates, the difference in masculinity scores between masculine and feminine clusters (Masculinity Bias, 1) was in a direction that was unexpected from either prediction: most scores fell significantly below 0. That is, inter-language theory makes its prediction based on the assumption that a relationship exists between grammatical categories in L1 and conceptual categories in L2. According to this theory, then, we should see higher masculinity scores among masculine cluster items compared to both feminine and neuter cluster items. On the other hand, GCM theory, taking a non-representational stance, predicts that there
would be no differences in masculinity scores between clusters. Indeed, there was no significant difference in masculinity scores between masculine and neuter clusters but, unexpectedly, feminine clusters outscored masculine clusters on masculinity. Thus, the apparent “reverse” masculinity bias effect was indeed unanticipated because neither theory provided a theoretical basis for this finding (for more discussion, see Exploring Reverse Masculinity Bias).

**Testing Intra-speaker Theory**

A mixed three by two ANOVA was performed to examine the relation between L1 grammatical gender and L2 proficiency for evidence of masculinity bias among bilinguals of different levels of English proficiency. L1 gender was the within-subject factor with three levels, masculine, feminine, and neuter; L2 proficiency was the between-subject factor with two levels, low and high L2 content. There was a main effect of L1 grammatical gender, $F(2, 86) = 47.15, p < .01, \eta^2 = .53$, indicating that the three gender-based clusters differed in their masculinity scores; furthermore, L1 grammar accounted for 53% of the variance in the dependent variable, masculinity scores. There was no main effect of L2 content, $F(1, 43) = 34.15, p = .123$, indicating that L2 proficiency as measured by L2 content did not impact masculinity scores.

Finally, there was not a significant interaction between L1 grammatical gender and L2 content, $F(1, 43) = 22.80, p = .295$.

A post-hoc test using the Bonferroni correction$^{20}$ (Glass & Hopkins, 1996) was conducted to examine which clusters scored higher on masculinity: the mean difference of 9.56 (95% CI, 6.96 to 12.16) between feminine and masculine and the mean difference of 7.69 (95% CI, 5.32 to 10.06) between feminine and neuter clusters were statistically significant at the $p < .05$ level. There was no statistically significant difference between neuter and masculine clusters.

$^{20}$The Bonferroni correction is used to correct or adjust the original level of alpha when multiple tests are conducted. That is, the original value of alpha is divided by the number of tests performed. The resulting corrected alpha is more conservative.
Overall, the results indicated that for feminine-cluster nouns, bilinguals tend to produce adjectives that scored higher on masculinity compared to either masculine- or neuter-cluster nouns.

The above findings provided no evidence in support of intra-speaker theory: there was no significant interaction effect between L1 grammatical gender and L2 proficiency. The only significant finding was not part of the alternative hypothesis since it concerned a main effect, i.e., feminine clusters’ scoring higher on masculinity than either masculine or neuter clusters. These results mirrored the findings about an apparent “reverse” masculinity bias effect discussed in the previous section (see Testing Inter-language Theory).

Exploring “Reverse” Masculinity Bias

The unexpected finding that feminine-cluster adjectives tend to score higher on masculinity than masculine-cluster adjectives, and that the trend persists across levels of L2 content prompted a closer look at the adjectives generated by the study’s participants. To that end, the pooled list of adjectives \((N = 62)\) was inspected for over- or under-representation of either low- or high-masculinity adjectives and then also compared to a corpus-based dataset to examine whether more masculine adjectives (e.g., with scores of 7 or 6) were overrepresented in either or both datasets.

First, a chi-square test of goodness-of-fit was performed to examine if more and less masculine adjectives were equally represented on the original list of 62 adjectives. The results were not statistically significant, \(X^2 (7, N = 62) = 8.19, p = .316\), indicating that there was no evidence of either low- or high-masculinity adjectives being over- or underrepresented. In other words, the list comprised adjectives of different types, that were equally represented across the masculinity spectrum.
Next, for each adjective, the number of times it was used by the study’s participants, i.e., its standard frequency count, was calculated. To test for the relation between masculinity and standard frequency based on the study’s sample, a correlation test was performed. (Upon examination for outliers, the adjective “green” was excluded from the analysis because its frequency value was 5.43 standard deviations greater than the sample’s mean of 8.13; it appeared 72 times in a sample size of 504. Standard frequency values ranged from 1 to 73.) There was a small positive relationship between masculinity and standard frequency, \( r(61) = .26, 95\% CI [0.02, .05], p = .04 \), suggesting that adjectives that scored higher on masculinity were used somewhat more frequently by the bilinguals compared to the adjectives with lower masculinity scores.

To examine whether the patterns observed in the study would be similar to those observed among a larger population of English speakers, a corpus-based dataset was consulted: SUBTLEX-US is an open-source database and search engine for the study of word frequency based on American subtitles (http://expsy.ugent.be/subtlexus/). But first, a log transformation was applied to the frequency data from the study as is customary in word frequency research\(^{21}\). The transformed set was then compared against the SUBTLEX-US dataset and each adjective was assigned the corresponding (log) frequency score from the corpus-based set. The two frequency measures, study- and corpus-based, were examined for correlation among themselves and with masculinity: the two frequency measures were moderately correlated, \( r(62) = .47, 95\% CI [.25, .64], p < .01 \), indicating that the study’s participants used various types of adjectives in a way that resembles that of English speakers in general. There was not a statistically significant relationship between corpus frequency and masculinity suggesting that English speakers do not

---

\(^{21}\) Studies that deal with word frequency often face a lot of unevenness and variation in individual data’s absolute frequency values (e.g., Criss, Aue, & Smith, 2011; Poarch & van Hell, 2012). Applying a log transformation to a dataset corrects for outliers as it makes the data less spread-out and its extreme values less extreme. In this case, the transformation allowed me to keep all 62 items for the subsequent analysis.
show a preference for either low or high masculinity adjectives. On the other hand, the
relationship between study-based frequency and masculinity was a small positive one, \( r(62) = .24, 95\% CI [-.01, .46], p = .064 \). This finding seemed to converge with the results of the earlier
correlation test between count frequency and masculinity, which found that the bilinguals tended
to use higher-masculinity adjectives more often than low-masculinity adjectives. It is important
to mention that, since the dataset of this study is relatively small by the standards of corpus-
based research, what appears as a significant result here may disappear once the size of the set
has increased.

Having ascertained that the patterns of adjective use in the study were consistent with that
of the general population and that there was no evidence of higher incidence of either more or
less masculine adjectives in the corpus, I then decided to re-examine the masculinity bias 1 data.
I hypothesized that with more English use bilinguals should show a more even distribution of
masculinity scores across clusters and thus demonstrate lower levels of reverse masculinity bias.
Since the test of intra-speaker theory showed no interaction effect between L2 proficiency and
L1 grammatical gender (masculinity bias), it seemed warranted to bring in an additional factor,
informal English: the two groups of speakers, low and high L2 content, differed in their use of
English language in informal settings, \( t(43) = -2.78, p < .01 \).

A two-by-two ANOVA was conducted to test for the effects of L2 content, informal
English and their interaction on the difference between masculine and feminine clusters
(masculinity bias). The overall model produced an \( F \) of 2.69, \( p = .058, \eta^2 = .17 \), indicating that
this model explains 17\% of the variability in the dependent variable, i.e., masculinity bias-I
values. (The same model, when performed without an outlier that scored 2.9 \( SD \) below the mean,

---

22 The obtained lower limit of -.01 is very close to 0, which suggests that the coefficient might have “dipped” below
zero because of the small sample size, especially since the strength of the relationship is also small (hence, the
obtained \( p \)-value of .064),
yielded a statistically significant overall $F$ of 3.56, $p = .022$; the rest of the results were in the
same direction as in the present model). There was a main effect of informal English, $F(1, 41) = 5.31, p = .026$: bilinguals, who used informal English more often, showed significantly less
evidence of “reverse” masculinity bias ($M = -7.36, SD = 7.55$) than those bilinguals who used
informal English infrequently ($M = -11.91, SD = 5.59$). Regarding effect size, the main effect of
informal English resulted in an $\eta^2$ of .12 (which is considered a medium to large effect size\(^{23}\)),
indicating that 12% of the variability in masculinity bias I values was explained by participants’
use of informal English. There was no main effect of L2 content, $F(1, 41) = 0.94, p=.338,$
meaning that L2 proficiency did not have an effect on the kind of adjectives bilinguals were
producing. There was a nonsignificant interaction effect, $F(1, 41) = 1.55, p = .220$.

These results suggest that those bilinguals who use English in informal settings often,
tend to follow native speakers’ patterns (e.g., use about the same number of more and less
masculine adjectives) and thus show significantly less evidence of “reverse” masculinity bias.
Estimated marginal means for the four groups are provided in Table 9 below.

\(^{23}\)Re-writing the resulting $\eta^2$ value of .12 as a Cohen’s $d$ produces a $d$-value of 0.75, which is a medium-to-large
effect size (Cohen, 1992). It must be noted, that although $d$-values are useful because they provide a unified
(standardized) way of comparing results, $d$-values tend to be biased when based on small samples (Cumming, 2014).
Table 9

*Estimated Marginal Means (M) for the “Reverse” Masculinity Bias Effect*

<table>
<thead>
<tr>
<th>Informal English</th>
<th>L2 Content</th>
<th>M</th>
<th>SE</th>
<th>95% CI</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>-11.64</td>
<td>1.99</td>
<td>-15.67</td>
<td>-7.60</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>-12.18</td>
<td>1.99</td>
<td>-16.22</td>
<td>-8.15</td>
</tr>
<tr>
<td>High</td>
<td>Low</td>
<td>-9.55</td>
<td>1.99</td>
<td>-13.58</td>
<td>-5.51</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>-5.17</td>
<td>1.91</td>
<td>-9.03</td>
<td>-1.31</td>
</tr>
</tbody>
</table>

*Note.* CI = confidence interval; LL = lower limit; UL = upper limit.

As summarized in Table 9, low- and high-L2 content bilinguals who used informal English infrequently had mean masculinity bias scores of -11.63 and -12.18, respectively. On the same measure, low-and high-L2 content bilinguals who used informal English frequently were found to have means of -9.55 and -5.17, respectively. Generally, as participants’ informal English scores went higher, there was less evidence of the “reverse” masculinity bias in their performance. What is more, higher informal English scores coupled with higher L2 content scores resulted in the lowest value of -5.17.

Overall it appears that in their use of adjectives – whether more or less masculine – bilinguals follow native speakers’ patterns; also, the more bilinguals use English in informal settings, the more their patterns resemble those of native speakers. The observed grouping of higher-masculinity adjectives around a particular cluster of nouns (aka “reverse” masculinity bias) may have to do more with the bilingual speaker’s exposure to and use of everyday informal English than with either L1 grammar or L2 proficiency based on formal measures of language.
Discussion

The main purpose of this study was to examine several interpretations of LR: intra-language, intra-speaker, and general cognitive mechanism theories. To test all three theoretical frameworks, I used a word association task, a popular means of examining influences of L1 grammar within the LR framework and of studying bilingual development (see Meara, 2002). Overall, I found evidence of language experience's influencing bilingual performance under L2 conditions.

Inter-language theory was not supported. In contrast to the received view (Boroditsky & Schmidt, 2000) that bilinguals of gendered languages associate certain English nouns with masculine adjectives following their native grammar (masculinity bias), I found no evidence of masculinity bias. The difference in masculinity scores between masculine and neuter clusters was not statistically significant, $t(44) = -1.58, p = .122$. Unexpectedly, I found evidence of reverse masculinity bias whereby feminine clusters received higher masculinity scores compared to masculine clusters, $t(44) = -9.14, p < .01$ (more discussion follows). The inconsistency with previous research may have to do with the spoken, online format of the used task. In contrast, Boroditsky and Schmidt (2000) relied on offline, written responses. However, if the Habitual Whorfianism view is correct, then spontaneous, spoken responses should be able to capture “habitual” behavior which did not happen in this study.

The unexpected finding of reverse masculinity bias converges with previous findings by Sera and colleagues (2002) who reported a considerable amount of contradiction in research on
the effects of grammatical gender on classification behavior among speakers of gendered languages: there are as many conflicting findings as there are gendered languages\textsuperscript{24}.

Intra-speaker theory did not receive support either. Contrary to Athanasopoulos and Kasai’s (2011) findings, less and more proficient bilinguals did not differ in their performance. Both groups followed the same pattern whereby feminine clusters scored higher on masculinity than either masculine or neuter clusters, with no difference between masculine and neuter clusters.

In a related line of studies on code-switching between a gendered L1 and a non-gendered L2, researchers have noticed a trend toward an overrepresentation of masculine gender so that some have suggested that masculine gender is the default choice in many gendered-L1/non-gendered L2 pairings. (In this context, code-switching usually refers to the practice of embedding isolated L2 words into L1 phrases.) There has been some disagreement among researchers about what underlies this process: whether bilinguals use semantics or formal (e.g., phonological) features to guide their gender assignment of the code-switched non-gendered L2 words. Violin-Wigent (2006) has reported that English nouns (non-gendered L2) when code-switched into French (gendered L1) tend to be ascribed a masculine article by French-English bilinguals. According to her, French-English bilinguals rely on semantically driven strategies as their preferred strategies for gender assignment (she did not consider whether relying on phonological features would have provided an alternative explanation). In contrast, Chirsheva (2009) has found that Russian-English bilinguals show a similar trend of masculine gender overrepresentation but the trend was based on formal (phonological) characteristics. In particular, Russian gender assignment tends to be based on such formal features as endings, e.g.,

\textsuperscript{24} For a very comprehensive and accessible treatment of gender as a grammatical category, see Corbett (2009). In many languages, grammatical gender is not a stable category. For example, phonetical changes could lead to the attrition of gender categories if two or more gender markings become phonetically undistinguishable.
in Russian a majority of nouns ending in a consonant are masculine; nouns ending in an unstressed vowel are usually feminine. Since many every-day common English nouns end in a consonant, masculine gender was found to be overrepresented among code-switching Russian-English bilinguals. There are English nouns that end in an unstressed vowel, but these nouns are relatively rare, usually borrowed words, such as *piñata* or *pizza*; those kinds of nouns were virtually absent in the study. Still, Chirsheva’s findings suggest that there are subtle L1 influences affecting gender assignment under code-switching conditions, influences that reflect L1 phonological speech patterns, rather than L1 grammar or semantics. It could be that a phonological factor, reflective of L1 speech patterns, affected the outcomes of the present study. Future research should consider using language experience measures that address both practices, L1 and L2, in bilinguals.

Finally, general cognitive mechanism theory received partial support. As predicted, the two groups of bilinguals, low and high L2 proficiency measured as L2 content, did not differ on their masculinity bias measures. However, contrary to the expectation, there was a significant difference in masculinity scores across clusters for both groups of bilinguals: both groups showed evidence of “reverse” masculinity bias. Additional analyses revealed that the patterns of adjective use among the study’s participants were similar to those of native English speakers. That prompted me to consider whether informal English could be a factor. A two-by-two ANOVA with L2 content and informal English as the two factors, returned a single significant finding – a main effect of informal English, $F(1, 41) = 5.31, p = .026$. This result suggests that the unexpected reverse masculinity bias effect was a function of language experience and use: bilinguals who used informal English frequently tended to resemble native English speakers’ responses to a greater degree and showed less reverse masculinity bias than those bilinguals who
used informal English infrequently. These results are in line with previous findings that with more L2 use, bilinguals’ performance tends to resemble that of native L2 speakers; and that L2 use correlates with task outcomes while formal measures of L2, such as L2 vocabulary, do not (Athanasopoulos, Damjanovic, Krajićova, & Sasaki, 2011).

This finding also expands on Kousta and colleagues’ (2008) study: they found that highly proficient early-onset Italian-English bilinguals, who reported preferential use of English over Italian, followed English native patterns when tested in English and showed no evidence of L1 grammar’s influencing their performance in English. In the present study, bilinguals spoke a different gendered language – Russian – as their first language; all were late-onset bilinguals. Despite these differences, I too found no evidence of L1 gender having an effect on bilinguals’ responses. Additionally, I observed a trend similar to Kousta and colleagues’ findings: in my study, bilinguals, who used informal English often, tended to produce responses that were similar to that of native English speakers.

Also, this finding lends additional support to the existing body of research on the importance of experiential factors in bilingual development vis-à-vis biological or innate factors. For a long time the field of bilingual research was under the strong hold of a maturational paradigm and its assumption that biological forces determined linguistic development among bilinguals. Similar to language development in children, it was assumed that once certain early neurological milestones were met, adult speakers were not expected to master a second language to the level of native speakers. One very influential study (Johnson & Newport, 1989) found that among early-onset bilinguals (age of onset < 16) there was a negative correlation between age

---

25 Kousta et al. (2008) did not make factors that related to language experience a focus of their study; their theoretical framework was representational and produced traditional psycholinguistic analysis.
and English proficiency; there was no relation between age and proficiency among late-onset bilinguals.

Later studies have challenged Johnson and Newport’s conclusions. Birdsong and Molis (2001) replicated the Johnson and Newport study and found the opposite pattern: they found no relation between age and proficiency among early onset bilinguals, and a negative correlation between age and proficiency among late-onset bilinguals. In addition, they found a positive relationship between amount of L2 use and L2 proficiency among late-onset bilinguals. Their findings received support from Moyer’s (2011) research who found age not to be a significant predictor of L2 proficiency. On the other hand, length of residence and L2 use (informal, with friends who are native speakers), or *quantity* and *quality* of L2 use respectively, were found to be significant.

The present study expands this line of research further: converging with Moyer’s (2011) findings, there was a significant difference in the amount of informal English use between low and high L2 content bilinguals (the two proficiency groups in the study). In addition, informal English was predictive of bilingual performance while L2 content (L2 proficiency) was not. This finding suggests that traditional linguistic measures of bilingual development such as L2 vocabulary and L2 syntax should be supplemented with experiential measures that capture speakers’ ongoing engagement with others, or in Moyer’s words, language experience (e.g., amount of L2 use in informal settings, types of informal settings, etc.).

Overall, the present study questions one’s overreliance on representational approaches: when we reduce language to a-temporal universals, including grammar, and then expect language to explain cognitive outcomes (e.g., classification or conceptual associations), the
results are inconsistent and less than satisfying. At the same time, the present findings call our attention to the fact that language serves social and communicative functions, and it is its use that engenders change and development in individual speakers and communities. We find early signs of this experience and usage-oriented view of language in Whorf’s original writing. Expounding on his notion of language, he asserted that "we cut nature up, organize it into concepts, and ascribe significance as we do, largely because we are parties to an agreement to organize it this way -- an agreement that holds throughout our speech community and is codified in the patterns of our language" (Whorf, 1956, p. 213).

Following this view of language as an enacted, implicit agreement that binds humans together, Nelson (2007) has proposed that language operates through shared cultural practices and thus provides young children with opportunity to join "a community of minds". Understood as ongoing engagement, language continues to exert its influence on speakers throughout their lives: it is through language that people find their place within historical and cultural settings as they partake in joint action with other speakers (Brockmeier, 2012).

This nonrepresentational view of language as shared experience prompts us to expand the study's findings further: what is it about language experience, including language use in informal settings, that makes it a significant factor in bilingual performance vis-à-vis traditional measures of L2 proficiency? Is it about the uneven fragmented (Brockmeier, 2012) ways in which speakers interact, especially in everyday casual exchanges? To what degree is language experience associated with changes and development in syntax, vocabulary or registers? All in all, it seems that we will have to consider a combination of linguistic, non-verbal and social and historical factors. This ambitious project would require a collective effort of psycholinguistics, cognitive

---

26 At best, we can expect to document cross-cultural differences, which are often difficult to replicate (e.g., January & Kako, 2007; Tse & Altarriba, 2008) and tend to disappear once we employ culturally sensitive measures (Cole, 1990).
science, and cultural psychology\textsuperscript{27}, if we want to advance our understanding of the role of language in human development and remain relevant to the needs of our community.

\textsuperscript{27} This interdisciplinary approach has been shown to be effective in bridging the gap between formal and experiential studies of language. For example, Ford, Fox and Thompson (2002) studied grammar as “woven” into everyday actions and were able to demonstrate how different syntactic structures help speakers accomplish various communicative goals.
CONCLUSION

The present studies sought to examine three interpretations of linguistic relativity using two popular tasks with a sample of late-onset Russian-English bilinguals. Since the area of LR research has grown to include various disciplines, with some considerable overlap in methods, I used two word association tasks and thus surveyed a broader area than I would otherwise if I had ran only one study.

Study one looked at the relation between language and cognition through the purported impact of L1 vocabulary on conceptual associations, study two – L1 grammar on associations. In both studies, there was no evidence in support of either representational view of LR, inter-language or intra-speaker. There was no evidence of either L1 vocabulary or L1 grammar’s influencing bilinguals’ conceptual associations; or bilinguals’ following L2 patterns as a function of L2 proficiency.

In contrast, both studies provided support for the non-representational reading of LR: the way bilinguals use L2 (e.g., with friends and family, in informal settings) impacts their performance under L2 conditions. In particular, more frequent use of informal English was associated with more consistent word associations in both languages (study one), and with using adjectives in a way similar to that of native English speakers (study two). Overall these findings call our attention to the importance of including language experience, or the multiple ways bilinguals use L2, in our studies of bilingual development. Equally important, these findings highlight the value of non-representational approaches to the study of language as not reducible to linguistic concepts or a universal grammar. To understand the impact of language, it should be studied as a language game, “consisting of language and the actions into which it is woven” (Wittgenstein, 2009/1953, §7, as cited in Brockmeier, 2012, p. 448). This very ambitious project
will require all the tools that we already have in our possession: theory and empirical findings of linguistics, cognitive psychology, and second language research. All this wealth of knowledge will need to be re-examined and further developed through a lens of cultural psychology. That way we will avoid committing an ontological fallacy (Brockmeier & Harré, 2001) as we refuse to reify abstract concepts and instead keep our research grounded in ongoing social interaction and cultural practice. Already we have seen how treating categories of Latin grammar as universal, in a context of languages other than English, has resulted in inconclusive and often conflicting findings. We should not be surprised by this, given that “our Greco-Roman pigeonholes … were not created with those languages in mind” (Tomasello, 2003, p. 18). Non-representational, socio-cultural approaches, on the other hand, allow for the inclusion of rich cultural context: after all, “verbal communication occurs contemporaneously with and not independently of other material and symbolic activities’ (Brockmeier & Harré, 2001, p. 42). That is, looking at language as embedded in social and cultural practice is a necessary next step if we are to understand language’s role in human development.

Moreover, by studying informal English alongside formal measures of L2 proficiency, the two studies attempted to bridge a gap between several research traditions, such as cognitive research and SLA studies: the former focuses more on linguistic characteristics of language and the latter on experiential. Indeed, focusing on informal English has provided additional evidence of the role of language experience in bilingual development as well as some ideas about how SLA happens in the real world: it appears that socially-meaningful everyday practice shapes bilingual development and has the potential to predict their performance in tasks that are usually measured in formal linguistic terms. In contrast, when bilingual development is studied as L2
vocabulary growth or the acquisition of L2 syntax, the result is often a static picture that provides a more or less clear distinction between novice and advanced bilinguals.

With the introduction of informal English, an experiential factor, we can now focus on identifying processes underlying bilingual development. One promising line of research involves longitudinal studies similar to developmental research with young children, which has provided a wealth of information and highlighted the importance of everyday language interaction between young children and their parents (Hart & Riley, 1999). Also, with the recent development of reliable statistical tools for use in single-case research design (Shadish, 2014), in-depth case studies of bilinguals can provide important insights into the nature of language, culture and cognition. Already we know that motivational factors contribute to bilingual development. For example, Ioup, Boustagui, Tigi, and Moselle (1994) have found that a speaker’s desire to become a full-fledged member of the host culture is associated with native-like L2 proficiency. Single-case studies have the potential to provide a comprehensive picture of bilingual development by combining self-reported, observational, and experimental measures over a period of time.

All in all, by expanding our understanding of bilingual development beyond the traditional dichotomy between linguistics-based and applied research, we can support the bilingual speaker and, equally important, learn more about the human potential in general.
# Appendix A: Screening Form

(Administered in Russian)

<table>
<thead>
<tr>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Source (GC mailing, etc.)</td>
</tr>
<tr>
<td>Contact information (phone or email)</td>
</tr>
</tbody>
</table>

## Notes:

<table>
<thead>
<tr>
<th>Current Age</th>
<th>Age at time of arrival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current employment/ school status, e.g., employed, full-time student, etc.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Use of Russian/ English: reports distinctive patterns of use, e.g. speaks Russian at home, English at work, in school, etc.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Reading skills – functionally literate: can read everyday print in Russian and English</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>History of reading/ learning disabilities: has never been diagnosed as RD or LD in Russia or US</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Qualifies to participate in the study:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Yes ☐ No ☐</th>
<th>Yes ☐ No ☐</th>
<th>Yes ☐ No ☐</th>
<th>Yes ☐ No ☐</th>
<th>Yes ☐ No ☐</th>
</tr>
</thead>
</table>
Appendix B: Consent Form

My name is Natalya Petroff and I am a student in the Educational Psychology Ph.D. Program at The Graduate Center of the City University of New York (CUNY), and Principal Investigator of this project, entitled “Reexamining linguistic relativity: What adult bilinguals can teach us about culture, language, and cognition”. This is a research study of bilingual competence. The study is expected to expand our knowledge about adult bilingual competence and contribute to effective second-language instruction.

I will ask you to perform two computer-based word-association tests, fill out a language history questionnaire, and take a brief fluency test, administered over the phone. The first word-association test will be conducted in Russian, and the rest of the study in English. The fluency test, Versant English, is administered and scored by Pearson, a commercial provider of testing and educational materials; to ensure your privacy, you’ll be assigned a number and only my advisor, I, and qualified Pearson personnel will have access to your audio file for research purposes only.

This study will take about one hour. You will receive a book-store gift card ($20) at the end of the study. With your permission, I would like to audio-record our session so that I can study your verbal responses in more detail later. The audio files will only be heard by me and my advisor, Dr. Bruce Homer. All information gathered will be kept strictly confidential, and will be stored in a locked file cabinet, to which only I and my advisor will have access. At any time you can refuse to answer any questions or end this interview.

This study involves no known risks. The benefit of your participation is that we expect that this study will add to what we already know about second language acquisition in adults and
help us develop effective second language instruction. In addition, you’ll be receiving a printed report, based on your fluency test, outlining your strengths in English. There will be approximately 40 participants taking part in this study.

I may publish results of the study, but names of people, or any identifying characteristics, will not be used in any of the publications. If you would like a copy of the study, please provide me with your address and I will send you a copy in the future (address/ email: _____________________________). Also, if you would like to be contacted about future research projects, please indicate here: yes □, no □.

If you have any questions about this research, you can contact me at 212-817-8298 or npetroff@gc.cuny.edu, or my advisor Dr. Bruce Homer at (212) 817-8292 or bhomer@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you for your participation in the study. I will give you a copy of this form to take with you.

<table>
<thead>
<tr>
<th>I agree to have this study session audio-recorded.</th>
<th>Please circle one:</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant Signature</th>
<th>Date</th>
<th>Investigator Signature</th>
<th>Date</th>
</tr>
</thead>
</table>
# Appendix C: Stimulus Words

## Stimulus Words

<table>
<thead>
<tr>
<th></th>
<th>Russian</th>
<th>English</th>
<th>Part of Speech</th>
<th>Concrete/Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ЯБЛОКО</td>
<td>APPLE</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>2</td>
<td>СПАЛЬНЯ</td>
<td>BEDROOM</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>3</td>
<td>КУХНЯ</td>
<td>KITCHEN</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>4</td>
<td>ДВОРЕЦ</td>
<td>PALACE</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>5</td>
<td>МОРЯК</td>
<td>SAILOR</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>6</td>
<td>УЧИТЕЛЬ</td>
<td>TEACHER</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>7</td>
<td>ДЯДЯ</td>
<td>UNCLE</td>
<td>Noun</td>
<td>Concrete</td>
</tr>
<tr>
<td>8</td>
<td>ОПАСНОСТЬ</td>
<td>DANGER</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>9</td>
<td>ТЕМНОТА</td>
<td>DARKNESS</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>10</td>
<td>СЕСТРА</td>
<td>SISTER</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>11</td>
<td>УСПЕХ</td>
<td>SUCCESS</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>12</td>
<td>КРАСОТА</td>
<td>BEAUTY</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>13</td>
<td>СЛАБОСТЬ</td>
<td>WEAKNESS</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>14</td>
<td>МУДРОСТЬ</td>
<td>WISDOM</td>
<td>Noun</td>
<td>Abstract</td>
</tr>
<tr>
<td>15</td>
<td>ДОЧЬ</td>
<td>DAUGHTER</td>
<td>Noun, Adj.</td>
<td>Concrete</td>
</tr>
<tr>
<td>16</td>
<td>ПАЛЕЦ</td>
<td>FINGER</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>17</td>
<td>ЗЕРКАЛО</td>
<td>MIRROR</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>#</td>
<td>Russian</td>
<td>English</td>
<td>Part of Speech</td>
<td>Concrete/Abstract</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>-----------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>18.</td>
<td>МАТЬ</td>
<td>MOTHER</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>19.</td>
<td>КАРТИНА</td>
<td>PICTURE</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>20.</td>
<td>ТЮРЬМА</td>
<td>PRISON</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>21.</td>
<td>ВОДА</td>
<td>WATER</td>
<td>Noun, Verb</td>
<td>Concrete</td>
</tr>
<tr>
<td>22.</td>
<td>УСЛУГА</td>
<td>FAVOR</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>23.</td>
<td>ЧЕСТЬ</td>
<td>HONOR</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>24.</td>
<td>ПРЕДЕЛ</td>
<td>LIMIT</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>25.</td>
<td>УДОВОЛЬСТВИЕ</td>
<td>PLEASURE</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>26.</td>
<td>БЕЗОПАСНОСТЬ</td>
<td>SAFETY</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>27.</td>
<td>МОЛЧАНИЕ</td>
<td>SILENCE</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>28.</td>
<td>РАССТОЯНИЕ</td>
<td>DISTANCE</td>
<td>Noun, Verb</td>
<td>Abstract</td>
</tr>
<tr>
<td>29.</td>
<td>СЕРДИТЫЙ</td>
<td>ANGRY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>30.</td>
<td>ГОРЬКИЙ</td>
<td>BITTER</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>31.</td>
<td>СУМАСШЕДШИЙ</td>
<td>CRAZY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>32.</td>
<td>ГРЯЗНЫЙ</td>
<td>DIRTY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>33.</td>
<td>СМЕШНОЙ</td>
<td>FUNNY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>34.</td>
<td>ЗДОРОВЫЙ</td>
<td>HEALTHY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
<tr>
<td>35.</td>
<td>СЧАСТЛИВЫЙ</td>
<td>HAPPY</td>
<td>Adjective</td>
<td>Abstract</td>
</tr>
</tbody>
</table>
Appendix D: Language History Questionnaire

Language History Questionnaire

Initials: ______________________ Date: ______________________

Instructions: Please answer the following questions to the best of your knowledge. In all your ratings and time estimations, don’t try to provide exact numbers, rather rely on your best guess/ approximation. When in doubt, go with your first hunch.

PART A

1. Age (in years): ______________

2. Sex (circle one): Male / Female

3. Education (highest degree obtained or highest school level attended):

_______________________________________________________________________

4(a) Country of birth: _____________________________________________________

4(b) Country of residence: _________________________________________________

5. How long have you lived in the United States (in years)? _________________

6. What is your native language? __________________________________________

7. Do you speak a second language?

   ___YES     My second language is ________________________________

   ___NO     (If you answered NO, you need not to continue this form)

8. If you answered YES to question 7, please specify the age at which you started to learn your second language (English) in the following situations (write age next to any situation that applies).

   At home: at age __________;     In school: at age __________

   After arriving in the second language speaking country: at age __________
9. How did you learn your second language (English) up to this point? (circle all that apply)

(Mainly Mostly Occasionally) through formal classroom instruction.

(Mainly Mostly Occasionally) through interacting with people.

A mixture of both, but (More classroom More interaction Equally both).

Other (specify): ____________________________________________

10. Rate your ability on the following aspects in each language. Please rate according to the following 7-point scale (write down the number in the table): Very poor (1) Poor (2) Fair (3) Functional (4) Good (5) Very good (6) Native-like (7)

<table>
<thead>
<tr>
<th>Language</th>
<th>Reading proficiency</th>
<th>Writing proficiency</th>
<th>Speaking fluency</th>
<th>Listening ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other (specify):

Other (specify):

11. Provide the age at which you were first exposed to foreign languages (English and/or other) in terms of speaking, reading, and writing, and the number of years you have spent on learning each language:

<table>
<thead>
<tr>
<th>Language</th>
<th>Age first exposed to the language</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
<th>Number of years learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. Do you have a foreign accent in the languages you speak? If so, please rate the strength of your accent according to the following 6-point scale (write down the number in the table):

No Accent (1)  Very Weak (2)  Weak (3)  Intermediate (4)  Strong (5)  Very Strong (6)

<table>
<thead>
<tr>
<th>Language</th>
<th>Accent (circle one)</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>Y       N</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Y       N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y       N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y       N</td>
<td></td>
</tr>
</tbody>
</table>

PART B

13. Estimate, in terms of percentages, how often you use your native language and other languages per day (in all daily activities combined, circle one that applies):

Native, or first language (Russian):  <25%  25%  50%  75%  100%

Second language (English):  <25%  25%  50%  75%  100%

Other languages (specify: ____________):  <25%  25%  50%  75%  100%

14. Estimate, in terms of hours per day, how often you are engaged in the following activities with your native and second languages.

<table>
<thead>
<tr>
<th>Activities</th>
<th>Russian</th>
<th>English</th>
<th>Other (specify: _________)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen to Radio/Watching TV:</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Reading for fun:</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Reading for work:</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
<td>______(hrs)</td>
</tr>
</tbody>
</table>
Activities | Russian | English | Other (specify: _________)  
--- | --- | --- | ---  
Reading on the Internet: | ______(hrs) | ______(hrs) | ______(hrs)  
Writing emails to friends: | ______(hrs) | ______(hrs) | ______(hrs)  

15. Estimate, in terms of hours per day, how often you used to speak your native and second languages with the following people.

<table>
<thead>
<tr>
<th>Language</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse/Partner:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Parents:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Children:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Grandparents(s):</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Brother(s)/Sister(s):</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Other family members:</td>
<td>______(hrs)</td>
</tr>
</tbody>
</table>

16. Estimate, in terms of hours per day, how often you now speak your native and second languages with the following people.

<table>
<thead>
<tr>
<th>Language</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse/Partner:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Children:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Relatives:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Friends:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Classmates:</td>
<td>______(hrs)</td>
</tr>
<tr>
<td>Co-workers:</td>
<td>______(hrs)</td>
</tr>
</tbody>
</table>
17. Write down the name of the language in which you received instruction in school, for each schooling level:

   Primary/Elementary School: _________

   Secondary/Middle School: _________

   High School: _________

   College/University: _________

18. In which languages do you usually:

   Count, add, multiply, and do simple arithmetic? ________________

   Dream? ________________

   Express anger or affection? ________________

19. When you are speaking, do you ever mix words or sentences from the two or more languages you know? (If no, skip to question 21.) ______

20. List the languages that you mix and rate the frequency of mixing in normal conversation with the following people according to the following 5-point scale (write down the number in the table): Rarely (1) Occasionall (2) Sometimes (3) Frequently (4) Very Frequently (5)

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Languages mixed</th>
<th>Frequency of mixing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spouse/Family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-workers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classmates</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
21. In which language (among your best two languages) do you feel you usually do better? Write the name of the language under each condition.

<table>
<thead>
<tr>
<th>Activity</th>
<th>At home</th>
<th>At work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>Writing</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>Speaking</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>Understanding</td>
<td>_______</td>
<td>_______</td>
</tr>
</tbody>
</table>

22. Among the languages you know, which language is the one that you would prefer to use in these situations?

<table>
<thead>
<tr>
<th>Location</th>
<th>Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>At home</td>
<td>_______</td>
</tr>
<tr>
<td>At work</td>
<td>_______</td>
</tr>
<tr>
<td>At a party</td>
<td>_______</td>
</tr>
<tr>
<td>In general</td>
<td>_______</td>
</tr>
</tbody>
</table>

23. If you have lived or travelled in other countries for more than three months, please indicate the name(s) of the country or countries, your length of stay, and the language(s) you learned or tried to learn.

___________________________________________________________________________

24. If you have taken a standardized test of proficiency for languages other than your native language (e.g., TOEFL or Test of English as a Foreign Language), please indicate the scores you received for each:

<table>
<thead>
<tr>
<th>Language</th>
<th>Scores</th>
<th>Name of the Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
<tr>
<td>_______</td>
<td>_______</td>
<td>_______</td>
</tr>
</tbody>
</table>
25. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

_______________________________________________________________________

_______________________________________________________________________

PART C

(Do you have additional questions that you feel are not included above? If yes, please write down your questions and answers below or on separate sheets.)
Appendix E: Cue Words

Cue Words

<table>
<thead>
<tr>
<th>##</th>
<th>Cue</th>
<th>Grammatical Gender Assignment in Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Grass</td>
<td>Feminine</td>
</tr>
<tr>
<td>2.</td>
<td>Lightning</td>
<td>Feminine</td>
</tr>
<tr>
<td>3.</td>
<td>Mountain</td>
<td>Feminine</td>
</tr>
<tr>
<td>4.</td>
<td>Night</td>
<td>Feminine</td>
</tr>
<tr>
<td>5.</td>
<td>River</td>
<td>Feminine</td>
</tr>
<tr>
<td>6.</td>
<td>Brook</td>
<td>Masculine</td>
</tr>
<tr>
<td>7.</td>
<td>Day</td>
<td>Masculine</td>
</tr>
<tr>
<td>8.</td>
<td>Flower</td>
<td>Masculine</td>
</tr>
<tr>
<td>9.</td>
<td>Hill</td>
<td>Masculine</td>
</tr>
<tr>
<td>10.</td>
<td>Thunder</td>
<td>Masculine</td>
</tr>
<tr>
<td>11.</td>
<td>Cloud</td>
<td>Neuter</td>
</tr>
<tr>
<td>12.</td>
<td>Field</td>
<td>Neuter</td>
</tr>
<tr>
<td>13.</td>
<td>Morning</td>
<td>Neuter</td>
</tr>
<tr>
<td>14.</td>
<td>Sea</td>
<td>Neuter</td>
</tr>
<tr>
<td>15.</td>
<td>Tree</td>
<td>Neuter</td>
</tr>
</tbody>
</table>
Appendix F: Adjective Rating Form

Instructions: Please rate each adjective as referring to either masculine or feminine properties of objects. Use only one checkmark per adjective. Be spontaneous and quick in your answers – don’t over think it! When finished, proceed to page 2.

<table>
<thead>
<tr>
<th>##</th>
<th>Adjective</th>
<th>Feminine</th>
<th>Masculine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Please provide your demographics by checking off boxes below:

Age: 18-30 □ 31-40 □ 41-50 □ 51-60 □ 61-70 □

Gender: Female □ Male □

English is my first language: Yes □ No □

I am fluent in other language(s): Yes □ Language(s): ________________
                                      No □

Education Level Completed (check all that apply):

High School Diploma □

2-year College (Associate’s) □

4-year College (Bachelor’s) □

Master’s Degree □

PhD/Professional Degree □

If you would like to receive a copy of our findings, enter your email below:

__________________________________________________________________

THANK YOU!
## Appendix G: List of Pooled Adjectives

### Pooled Adjectives

<table>
<thead>
<tr>
<th>##</th>
<th>Adjective</th>
<th>##</th>
<th>Adjective</th>
<th>##</th>
<th>Adjective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Beautiful</td>
<td>22.</td>
<td>Fluffy</td>
<td>43.</td>
<td>Refreshing</td>
</tr>
<tr>
<td>2.</td>
<td>Big</td>
<td>23.</td>
<td>Funny</td>
<td>44.</td>
<td>Roaring</td>
</tr>
<tr>
<td>5.</td>
<td>Bright</td>
<td>26.</td>
<td>Great</td>
<td>47.</td>
<td>Scary</td>
</tr>
<tr>
<td>10.</td>
<td>Colorful</td>
<td>31.</td>
<td>Hectic</td>
<td>52.</td>
<td>Snowy</td>
</tr>
<tr>
<td>13.</td>
<td>Dangerous</td>
<td>34.</td>
<td>Late</td>
<td>55.</td>
<td>Stormy</td>
</tr>
<tr>
<td>15.</td>
<td>Deep</td>
<td>36.</td>
<td>Loud</td>
<td>57.</td>
<td>Sunny</td>
</tr>
<tr>
<td>17.</td>
<td>Easy</td>
<td>38.</td>
<td>Peaceful</td>
<td>59.</td>
<td>Tall</td>
</tr>
<tr>
<td>18.</td>
<td>Endless</td>
<td>39.</td>
<td>Pretty</td>
<td>60.</td>
<td>White</td>
</tr>
<tr>
<td>20.</td>
<td>Fabulous</td>
<td>41.</td>
<td>Rainy</td>
<td>62.</td>
<td>Young</td>
</tr>
<tr>
<td>21.</td>
<td>Fast</td>
<td>42.</td>
<td>Red</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


