12-14-2018 11:00 AM

The Effects of Proficiency and Task Context on L2-L1 Noncognate Masked Translation Priming in Chinese-English Bilinguals

Mark J. McPhedran, *The University of Western Ontario*

Supervisor: Lupker, Stephen J., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Psychology

Abstract

The masked translation priming effect was examined in Chinese-English bilinguals using three experimental paradigms: lexical decision, semantic categorization, and speeded episodic recognition. A machine-learning approach was used to assess the subject- and item-specific factors that contribute to the sizes of translation priming effects across these tasks. The factors that contributed to translation priming effects were found to be task-specific. Priming effects in lexical decision were associated with higher self-rated listening and writing abilities in English, especially when primes were high-frequency and targets were low-frequency. Priming effects in semantic categorization were associated with more frequent use of English in daily life, especially when targets were high-frequency and primes were low-frequency. Finally, priming effects in episodic recognition were associated with higher self-rated reading, writing, speaking, and listening abilities in English. These results are discussed within different frameworks of current models of bilingual language processing.

Keywords: Masked translation priming, bilingualism, lexical decision, semantic categorization, episodic recognition

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Dr. Stephen Lupker for all of the support and advice that he gave me over the course of my six years at Western, and over the course of the four years of my doctoral program. His experience, insight, and timely feedback were invaluable in helping me get through the process of writing this dissertation, and always kept me grounded and focused throughout this ordeal. Thank you for being such a fantastic mentor, and helping me with all of the timely edits that you gave me.

I would also like to thank my labmates, and those no longer with the lab, including: Dr. Jason Perry, Alex Taikh, Giacomo Spinelli, Huilan Yang, and Zian Chi for all of their encouragement and comments that they have given me over the course of writing this dissertation.

To my dissertation committee members, Dr. Debra Jared and Dr. Albert Katz, I would like to express my gratitude for many of the helpful suggestions and pointers that you have given me in preparing to conduct this research.

To my parents, Cynthia and Stephen McPhedran. Everything that I have worked towards these last 10 years I could not have done without you. Thank you for everything that you have done to support me through every step of my life, and for always nurturing that scientific curiosity in me every step of the way.

To my in-laws, Viji Balakrishnan and Bala Pandalangat. Thank you for always going out of your way to invest your time in me, providing me with emotional support and professional guidance, and for always trying to bring out the best in me throughout this entire process.

And last, but certainly not least, I would like to thank my wife, Anjana Balakrishnan. Through the thick and thin, the good, the bad, and the ugly, your support for me never wavered. You were a rock for me in accomplishing my goals, and none of this would have been possible without your undying ability to always keep me motivated. I could not ask for a better partner in life.

Table of Contents

List of Tables

List of Figures

Chapter 1

1 Introduction

It is now estimated that over half of the world's population of seven billion people speak more than one language (e.g., European Commission Special Eurobarometer, 2006, 2012), and nowhere is bilingualism more prevalent than in Europe, where it is now estimated that 19% of people are bilingual, 25% are trilingual, and 10% speak four or more languages. Being able to communicate in multiple languages directly affects the mobility of workers within the European Union. Thus, it is no surprise that the EU has been encouraging its constituent states to push policy objectives that seek to establish a trilingual population, where citizens would be educated in their native language, English, and one of the other 22 languages spoken in the EU. Even more relevant, perhaps, is the case of Canada, where both English and French have legal equality in Parliament as well as in the court systems, and where access to many jobs within the government requires the ability to provide services in both English and French. Reflecting this policy of official bilingualism is the fact that French second-language education is a core part of the school curriculum in most provinces.

From the perspective of cognitive psychology, one issue that having a bilingual curriculum raises is whether doing so affects students' ability to learn, or, more specifically, their cognitive development. Cognitive psychologists have spent decades debating whether exposing children to multiple languages affects children's development, and whether there are negative consequences of doing so. The most common assumption was that learning two languages would be confusing for children, and that their cognitive abilities would lag behind their monolingual peers (e.g., Hakuta, 1986), with studies showing that bilingual children and adults have smaller vocabulary sizes in each language than their monolingual counterparts (e.g., Bialystok, Luk, Peets, & Yang, 2010; Gollan, Montoya, & Werner, 2002), have sparser semantic representations for words in both languages than monolinguals (e.g., Verhallen & Schoonen, 1998), and show slower comprehension and production of words even in their dominant language (e.g., Ivanova & Costa, 2008; Randsell & Fischler, 1987). In contrast, other studies have shown that bilinguals demonstrate better executive control (e.g., Bialystok, Craik, & Luk, 2012; however, see Paap & Greenberg, 2013). One thing is clear from this research: learning a second language has a fundamental impact on one's cognitive development, and this impact can be both positive and negative.

The scope of bilingualism research extends beyond investigating the effects of learning a second language on one's executive functioning and language learning, however. To understand why these issues might arise in the first place, one must understand the effects of learning a second language on the organization of language representations in memory. Accounts of the effects of bilingualism on executive functioning, for example, often assume that any advantage for bilinguals stems from having to manage attention to two languages, and actively suppressing the activity of one language in memory to use the language that is appropriate in the current context of use (e.g., Green, 1998; Norman & Shallice, 1986). Such an account assumes that both languages are always activated, and that there is some level of interaction between them, even in monolingual contexts. Understanding the nature of how the two languages are connected and represented in memory, then, is a critical question that must be addressed.

1.1 Translation Priming Paradigms

Questions of how bilingual memory is organized have been typically answered using data from behavioural experiments. One of the most common experimental paradigms used is the translation priming paradigm. In this paradigm, a prime is presented in one language, followed by a target that is either a translation equivalent of the prime, or is unrelated to the prime (e.g., 国王 (king) → KING vs. 鹹肉 (bacon)→ KING), and the subject must then make a decision on the target, typically a word-nonword decision. The assumption behind using translation priming is that, if the two languages are interconnected within lexical and semantic memory, using primes that are translation equivalents of the targets should preactivate lexical and semantic information about the target, making decisions on the target faster than when such information is not preactivated.

In one of the earliest studies done on translation priming effects, Meyer and Ruddy (1974) had German-English bilinguals classify letter string pairs as either words (e.g., HORSE-ACHT) or nonwords (e.g., SLATSCH-PERSAGE) in the two languages. Meyer and Ruddy found that word pairs that were semantically associated with each other were classified more quickly than unassociated pairs, and the size of the effect was just as large when the paired words were from different languages (e.g., SIEBEN-EIGHT) as when the pairs were from the same language (e.g., SEVEN-EIGHT). Other early research showed that these apparent cross-language “priming” effects occur only when the target stimulus immediately follows the prime in the different- (i.e., between-) language condition. For example, Kirsner, Brown, Abrol, Chadha, and Sharma (1980) had Hindi-English bilinguals complete a lexical decision task, (i.e., subjects had to decide whether each individually presented target was a word or a nonword). The experiment consisted of two blocks. In the first block, subjects had to respond to targets that could be either English or Hindi words or nonwords. In the second block, the original words were either repeated in the same language, or in the other language, and these words were mixed in with new words and nonwords. Using this paradigm, Kirsner et al. found a benefit of repetition when the target was repeated in the same language, but found little to no facilitation when the repetition was between-languages. Based on these findings, Kirsner et al. argued for a language-specific view of bilingual lexical representation.

In a follow-up study using French-English bilinguals, however, Kirsner, Smith, Lockhart, King, and Jain (1984) found that between-language translation priming does occur when the target is presented immediately after its translation equivalent is presented in a more standard priming paradigm, and argued that these results mean that, while bilingual lexical representations are language-specific, the lexicons function within an integrated network. Other early work by Schwanenflugel and Rey (1986) extended the findings of Kirsner et al. using Spanish-English bilinguals. These experiments used short prime-target stimulus onset asynchronies (SOAs) of 100 ms and 300 ms. Schwanenflugel and Rey found that the priming effects for cross-language (i.e., translation) primes were no different than for same-language primes, regardless of the SOA, and interpreted these results as meaning that bilingual lexical representations are connected by a representational system that is independent of language. In the intervening years, studies have repeatedly shown that translation priming is inevitably found when subjects are given an appropriate amount of processing time, regardless of whether the languages have a common script (e.g., Frenck & Pynte, 1987; Grainger & Beauvillain, 1988) or used different scripts (e.g., Chen & Ng, 1989).

The present research, unlike much of the early research, did not use a visible priming paradigm, as there are limitations to the conclusions one can draw from visible priming paradigms. Perhaps the most obvious issue is that subjects are consciously aware of the prime's existence, and, as a result, can strategically use the prime to aid in making decisions about the target. For example, having a conscious appreciation of the prime can result in subjects generating expectations about what target will follow the prime, and using those expectations to prepare their response in advance. Such strategic processes may tell us little about the nature of bilingual lexical memory. Further, because the subject is fully aware that the task involves processing in their L2 and L1, subjects could then become aware of the purpose of the prime, which may induce a subject-expectancy effect that biases the results of the experiment. Thus, while evidence from tasks using visible primes can provide some insights into how bilinguals' lexical representations are organized in memory, a much stronger source of evidence would come from a paradigm that minimizes strategic processes, and which masks the bilingual nature of the task. Any results from such experiments can thus be thought of as providing a methodologically purer measure of bilingual lexical processing. The masked priming paradigm (Forster & Davis, 1984) was designed with this exact goal in mind.

Masked priming is an experimental paradigm that was developed by Forster and Davis (1984), in which a prime (e.g., the nonword *homse*) is presented for a very brief period of time (~50 ms), and is sandwiched between a forward mask (e.g., #####) and a target to which the subject must respond (e.g., *HOUSE*), typically by making a word-nonword decision. Because the prime is presented so briefly and both forward and backward masked, few, if any, subjects are aware of its identity or even of its existence. Therefore, it is normally assumed that priming effects obtained in the masked priming paradigm must be due to automatic processes, because subjects are not consciously aware of any relationships between the prime and target stimuli. Critically, even though the prime is unavailable to consciousness, this paradigm has been found to produce robust effects on target processing latencies. For example, the word *HOUSE* is recognized significantly faster when it is primed by an orthographically similar nonword such as *homse* than when it is primed by a control nonword prime such as *clinb*.

Based on a general acceptance of these assumptions concerning the masked priming paradigm, that paradigm has been frequently used in bilingualism research. As with the unmasked version of the translation priming task, the masked translation priming paradigm involves presenting a prime in one language, followed by a target which is either a translation equivalent of the prime, or an unrelated word (which is also in the other language). In the masked version of the task, however, the prime is presented for only a very brief duration (~50 ms), and is typically sandwiched between a forward mask (#####) and the target. If a bilingual's first (L1) and second (L2) languages share a common representation in memory, or, at the very least, the language representations interact with each other in memory, presenting a prime in one language (to be followed by its translation equivalent target in the other language) should preactivate the meaning of the target, making responses to those targets faster.

One of the first attempts to examine bilingual language processing using the masked translation priming paradigm was reported by de Groot and Nas (1991), who studied Dutch-English bilinguals using cognate and noncognate translation pairs. Cognates refer to translation equivalents that, typically, have the same origin, and, as a result, have similar spellings and/or pronunciations (e.g., *wife* and *wijf*), whereas noncognates refer to translation equivalents with different spellings and sound patterns in the two languages (e.g., *pants* and *broek*). In the cross-language priming conditions in their first two experiments, de Groot and Nas used cognate prime-target pairs in their translation condition, whereas noncognate prime-target pairs were used in the translation condition in Experiments 3 and 4. In Experiment 1, they presented prime-target pairs, which were either within-language (i.e., English-English, Dutch-Dutch) or cross-language (i.e., Dutch-English, English-Dutch), and were either repetition/translation prime-target pairs (e.g., ground-GROUND, grond-GROND, grond-GROUND, ground-GROND), associatively related (e.g., calf-COW, kalf-KOE, kalf-COW, calf-KOE), or unrelated (e.g., bride-TASK, bruid-TAAK, bruid-TASK, bride-TAAK). In addition to finding substantial priming effects for the cognate prime-target pairs, de Groot and Nas also found significant cross-language associative priming in both the L1-L2 (i.e., Dutch primes and English targets) and the L2-L1 (i.e., English primes and Dutch targets) direction. In their second experiment, de Groot and Nas (1991) successfully replicated those findings. That is, significant priming effects were found again for not only direct translation pairs (e.g., koe-COW), but also for associatively related cognate pairs (e.g., kalf-COW). In their third and fourth experiments, de Groot and Nas found that using noncognates still produced significant masked translation priming effects in the L1-L2 direction, however, the priming effects for associative prime-target pairs disappeared. What de

Groot and Nas's research as well as results from subsequent studies have made clear is that between-language masked priming effects are contingent on several factors, including whether the prime-target pairs are cognates (e.g., Gollan, Forster, & Frost, 1997; Sanchez-Casas, Garcia-Albea & Davis, 1992, Experiment 1), and whether the prime-target pairs are direct translation equivalents of each other or are associatively related.

1.2 The Masked Priming Asymmetry

It should be noted that while de Groot and Nas (1991) studied masked cognate priming in both the L1-L2 and L2-L1 directions, their experiments using noncognates did not involve an L2-L1 condition. Note also that, given that cognates are visually and phonologically similar, cognates are likely to produce priming that goes beyond the priming due to the shared meaning of the words. Even in the case where there is no orthographic overlap between the two languages (e.g., English and Japanese), a shared sound pattern could also contribute to any cognate priming effect. The obvious question, therefore, is what is the nature of translation priming when noncognates, words that are not orthographically or phonologically similar, are used?

Whether masked translation priming would occur with noncognate prime target pairs in the L2-L1 direction was fully addressed by Gollan, Forster, and Frost (1997) using both Hebrew-English and English-Hebrew bilinguals. In each of their experiments, subjects were presented with English and Hebrew targets, which were either primed by within-language repetition and control primes (e.g., bunker-BUNKER vs. rodent-BUNKER; פִּירָמִידָה-פִּירָמִידָה vs. פִּירָמִידָה-רגליים) or by between-language translation (e.g., פִּירָמִידָה-PYRAMID, טִירָה-CASTLE) and control primes (e.g., רגליים-PYRAMID, סָגֹול-CASTLE). Both cognate and noncognate pairs were used. Primes were presented in the L1-L2 direction in their first two experiments, and in the L2-L1 direction in their last two experiments. As with de Groot and Nas (1991), Gollan et al. found significant masked translation priming effects for both cognates and noncognates when subjects were tested in the L1-L2 direction. Critically, however, Gollan et al. found that the priming effects, for cognates and noncognates alike, were eliminated when testing was done in the L2-L1 direction.

Similar results to Gollan et al.'s (1997) had been produced in previous unmasked priming tasks (e.g., Altarriba, 1991; Chen & Ng, 1989; Jin, 1990; Keatley, Spinks, & de Gelder, 1994).

Essentially, the clear trend observed across these experiments was that priming effects were larger when the prime was in the subject's L1 and the target was in the subject's L2. Even in Keatley et al.'s Experiment 3, where a significant L2-L1 priming effect was found, the priming effects for the L1-L2 direction were noticeably larger than the priming effects for the L2-L1 direction. More importantly, the asymmetric priming effects have been replicated multiple times over the last two decades and the most common finding in the literature has been that significant priming effects occur in the L1-L2 direction, while null priming effects are found in the L2-L1 direction (e.g., Chen, Zhou, Gao, & Dunlap, 2014; Dimitropoulou, Duñabeitia, & Carreiras, 2011a, 2011b; Finkbeiner, Forster, Nicol & Nakamura, 2004; Grainger & Frenck-Mestre, 1998; Jiang, 1999; Jiang & Forster, 2001; however, see Basnight-Brown & Altarriba, 2007; Duyck & Warlop, 2009; Nakayama, Ida, & Lupker, 2016; Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009).

1.3 Models of Bilingual Language Processing

1.3.1 The Episodic L2 Hypothesis

While it is clear from the research discussed above is that there is an asymmetry in the behavioural data that one obtains in translation priming, lexical decision tasks, with priming in the L2-L1 direction often not obtained, the debate over the theoretical mechanism that is responsible for producing this asymmetry remains unresolved. Several theoretical accounts have been proposed to account for the priming asymmetry. The first such theoretical account to be discussed is the Episodic L2 Hypothesis (Jiang & Forster, 2001).

The Episodic L2 Hypothesis is based on the idea that the reason one does not obtain L2-L1 translation priming effects in lexical decision is because L2 and L1 words are represented in different memory systems. Whereas L1 representations are assumed to reside in lexical memory, L2 representations are not. Rather, information about L2 words is assumed to be stored in episodic memory as a set of associations between L2 words and their L1 translation equivalents. That is, L2 information is represented episodically. This account argues that if the task is mediated by episodic memory processes, then an L2-L1 priming effect should be observed, whereas an L2-L1 priming effect should not be observed when the task is mediated by lexical memory processes because the representations of the L2 primes (being stored in episodic memory) would not activate the lexical representations of L1 words.

To test their account, Jiang and Forster (2001) used a masked L2-L1 translation priming paradigm in which subjects performed a speeded episodic recognition task. This task had two phases. In the first phase, subjects had to memorize a list of L1 words. In the second phase, subjects were presented with a mix of new words together with the old words, that is, the words that had previously been studied by the subject during the first phase of the task. Subjects had to decide whether each word was old or new as quickly and as accurately as possible. Most importantly, the words presented during the testing phase were primed by a masked prime in their L2. Jiang and Forster found significant L2-L1 masked translation priming in this task, however, crucially, the priming effect was only for words that had been previously presented during the training phase of the experiment (i.e., the "old" words, those that were stored in episodic memory). The priming effect for the new words was null. Further, using the same words that were presented in their speeded episodic recognition task, Jiang and Forster had subjects perform a masked L2-L1 translation priming task in which they had to make lexical decisions. As with prior research (e.g., Gollan et al., 1997; Grainger & Frenck-Mestre, 1998), the lexical decision task produced a null priming effect. Finally, Jiang and Forster had subjects perform the lexical decision task and the episodic recognition task in the L1-L2 direction. Under these circumstances, because the L1 words are represented in episodic memory, a null priming effect is predicted in the episodic recognition task, but a significant priming effect was predicted in the lexical decision task. Indeed, Jiang and Forster found that the episodic recognition task produced a null priming effect, while the lexical decision task produced a significant priming effect, consistent with the predictions of the Episodic L2 Hypothesis.

In subsequent research, Witzel and Forster (2012) further tested the Episodic L2 Hypothesis. In their first experiment, Witzel and Forster replicated Jiang and Forster's (2001) results that masked translation priming was produced in an episodic recognition task for studied L1 targets, but not for unstudied L1 targets, while at the same time replicating the asymmetry found in the lexical decision task (i.e., priming in the L1-L2 direction but not in the L2-L1 direction). In their second experiment, Witzel and Forster had subjects learn words in an unfamiliar language, and found that these words could prime their L1 translation equivalents in an episodic recognition task, but not in a lexical decision task. In a final experiment, Witzel and Forster examined masked repetition priming in an episodic recognition task. When English L1 speakers were tested, repetition priming (L1-L1) was found only for old words. However, when Chinese-English bilinguals were tested with the same items, a repetition priming effect was found for both old and new words. These results were interpreted as being consistent with the Episodic L2 Hypothesis, and as evidence that L2 words that are acquired later in life are represented in a different memory system than L1 words.

It must be pointed out that there is a serious problem for the Episodic L2 Hypothesis, however. That is, while this account can provide an adequate explanation of the task-specific differences between the episodic recognition task and the lexical decision task, Jiang and Forster's (2001) explanation has difficulty explaining the results from semantic categorization tasks (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), tasks that, as will be noted below, also show L2-L1 priming. If L2 words are unable to activate relevant lexical representations for L1 words because they are represented in a different memory system, then a task such as semantic categorization, which would require the activation of lexical representations in order to access semantic information, should also produce a null priming effect. Witzel and Forster (2012) attempted to address this issue by arguing that the episodic recognition task and the semantic categorization task have more in common with each other than with the lexical decision task, in that lexical decisions can be made without accessing meaning, while episodic- and semantic-based decisions cannot. However, even Witzel and Forster note that this argument runs into serious problems when one considers results in semantic priming experiments which show that semantic relationships are important in lexical decision tasks (see Neely, 1991, for a review), or results from semantic categorization tasks using broad or ad hoc categories which do not show L2-L1 priming, even though semantic activation is still clearly required (e.g., Wang & Forster, 2010).

Note also that, while it is entirely plausible that bilinguals' L2 information is initially represented in episodic memory, the Episodic L2 Hypothesis does not allow for the representations of L2 words to change over the course of L2 acquisition. It was instead assumed that the episodic links between L2 and L1 continue to be the sole relevant factor even for proficient L2 speakers. That proposition seems somewhat unrealistic for individuals who become quite proficient in their L2.

It is possible, however, that the Episodic L2 Hypothesis can account for how L2 words are represented within memory during the early stages of L2 acquisition, but over the course of becoming more proficient in their L2, the representations gradually migrate from episodic memory to lexical memory. Thus, the possibility that L2 representations migrate from episodic to lexical memory warranted examining.

1.3.2 The Distributed Conceptual Feature Model

The Distributed Conceptual Feature Model (DCFM; de Groot, 1992) provides another account of bilingual memory representation. This model assumes that bilinguals' L1 and L2 are represented by differentiated systems at the lexical level, but these differentiated systems are directly connected to each other. The model further assumes that the languages share a common conceptual system with a distributed, rather than localist, architecture. Words in L1 and L2 are, however, assumed to vary in how many of their features at the conceptual level overlap with each other. The more overlap at the conceptual level, the more semantically similar the two words are. This model is thus built on the idea that translation equivalents can have meanings that are language-specific, and will not overlap perfectly with each other.

The model makes what appears to be an easily testable assumption. It assumes that featural and conceptual overlap will depend on what type of word is represented. Therefore, translation priming effects would be larger for translation pairs that have more overlap in their conceptual representations. For example, as de Groot (1992, 1993) has argued, translation equivalents for concrete words should have more featural overlap than those for abstract words and, hence, should produce larger priming effects. Evidence concerning the viability of the DCFM (de Groot, 1992; de Groot, Dannenburg, & van Hell, 1994), therefore, comes from studies that have examined the effects of concreteness on translation priming. For example, in a study with Korean-English bilinguals, Jin (1990), using unmasked primes, found that concrete prime-target pairs produced larger priming effects than abstract prime-target pairs, regardless of whether the prime was a direct translation of the target, or was associatively related, supporting the model's prediction.

There are, however, several challenges for de Groot's (1992) model as well. First, the DCFM has difficulty accommodating the translation priming asymmetry. Regardless of translation direction, the model predicts equivalent priming effects, as the degree of featural overlap between the two words is constant regardless of prime-target direction. Further, while Jin's (1990) study found evidence of an interaction between prime type and concreteness, this interaction was specific to the L1-L2 direction. In the L2-L1 direction, the interaction between concreteness and priming effects disappeared for translation equivalents, although it remained for the associatively related prime-target pairs. Such a finding would appear to contradict the DCFM, as the translation equivalent prime-target pairs should still be assumed to have more featural overlap than the associatively related prime-target pairs, and should still yield larger priming effects for concrete words as a result.

A revised version of the DCFM (Kroll & de Groot, 1997) attempted to address the priming asymmetry problem by proposing that the connections between L2 lexical nodes and their conceptual features are weaker than the connections between L1 lexical nodes and their conceptual features for unbalanced bilinguals. Such a revision would, at least in theory, allow the model to account for the priming asymmetry in unbalanced bilinguals, while also accounting for why concreteness effects are weaker in the L2-L1 direction than the L1-L2 direction (see Jin, 1990; Schoonbaert et al., 2009, Experiments 1 & 2). One issue with this interpretation, however, is that this account would appear not to provide a mechanism that would explain the task-specific nature of the priming asymmetry effect, as subsequent research has shown that the priming effects obtained in the L2-L1 direction are sensitive to the nature of the target task (e.g., Finkbeiner et al., 2004; Jiang & Forster, 2001; Wang & Forster, 2010; Xia & Andrews, 2015). An account that is solely based on differences in connection strengths between L2 and L1 lexical nodes and conceptual features can plausibly predict weaker priming effects from L2 primes in any task, but still cannot explain why tasks such as semantic categorization and episodic recognition would produce an L2-L1 translation priming effect while a task such as lexical decision would not.

1.3.3 The Sense Model

Finkbeiner et al. (2004) proposed an alternative account of the priming asymmetry that was heavily based on the assumptions of the DCFM (de Groot, 1992; de Groot et al., 1994; Kroll & de Groot, 1997). Like the DCFM, Finkbeiner et al. assumed that lexical-level representations map onto distributed semantic representations. Where the Sense Model and the DCFM differ is that the Sense Model assumes that semantic representations are comprised of bundles of features bound together, corresponding to distinctive uses of each feature. They refer to these bundles of features as senses. Finkbeiner et al. largely base their ideas about semantic senses on research done by Rodd, Gaskell, and Marslen-Wilson (2002). According to Rodd et al., senses refer to systematic variations of a word's meaning according to the context in which it is used. As an example, Rodd et al. discusses how the word *twist* can have a variety of dictionary definitions, including "*to make into a coil or spiral to operate by turning*, *to alter the shape of it*, *to misconstrue the meaning of*, *to wrench or sprain*, and *to squirm or writhe*" (p. 245). Even though the meaning of the word varies due to the context, the interpretations of the word are closely related to each other.

Based on Rodd et al.'s (2002) account, Finkbeiner et al. (2004) argued that the semantic priming effect reflects the ability of prime words to preactivate semantic senses associated with the target words. The Sense Model is based largely on this idea, and makes a few key assumptions about the structure of these representations and, hence, about the nature of priming effects. First, it is assumed that words in both L1 and L2 are associated with several different senses, many of which are shared cross-linguistically. However, bilinguals who are acquiring their L2 may not be familiar with most of the senses associated with these words. Essentially, L1 words are associated with more semantic senses than their L2 translation equivalents. Second, and most importantly, it is assumed that the magnitude of priming produced by a prime is directly dependent on the number of senses that a prime can preactivate in a target. Priming can thus only occur in lexical decision tasks when primes are able to activate a sufficiently large proportion of the semantic senses that are associated with their targets. In the case of L1-L2 priming, when L1 primes are used, the senses that have been acquired for L2 words are more likely to be senses that are shared with their L1 translation equivalent. As a result, L1 primes preactivate a large proportion of the semantic senses associated with L2 targets, and a priming effect is observed.

On the other hand, when L2 primes and L1 targets are used, the L2 primes only preactivate a small subset of the semantic senses associated with the L1 targets. Thus, a null priming effect is observed.

The Sense Model makes several additional predictions. First and foremost, the sense model predicts that, even in monolingual tasks, masked priming effects should only occur when the prime contains virtually all the senses of the target, for example, when the prime contains many senses, and the semantically related target contains only one (shared) sense. Further, such a result should also be found in bilingual tasks, in that priming should only be obtained when targets with only a few senses that are known to the L2 learner and are shared with the prime are used. In contrast, even in the L1-L2 direction, using primes with a single sense and targets with multiple senses should produce a null priming effect.

Yet another interesting prediction made by the Sense Model is that the asymmetry should be sensitive to task context. Specifically, the asymmetry should not be produced in tasks in which the proportion of primed to unprimed senses is irrelevant to the decision in the task. Specifically, Finkbeiner et al. (2004) identified the semantic categorization task, where it is assumed that, while words may be associated with several different senses, the only senses that matter in such a task are the ones that contain category-relevant information. For example, English word black and the Japanese translation equivalent 黒い, while containing several senses that are language-specific and are not shared, contain the sense relevant for colour. In a semantic categorization task where subjects need to decide whether words are colours or not, only the sense that identifies the word as a colour is needed to make the decision and, hence, a translation priming effect would be expected in both directions.

Empirical support for the Sense Model is mixed. Evidence consistent with the Sense Model was reported by Grainger and Frenck-Mestre (1998). In their studies, Grainger and Frenck-Mestre had English-French bilinguals perform semantic categorization and lexical decision tasks with translation priming in the L2-L1 direction. In their experiments, primes were presented in French (subjects' L2), while targets were presented in English (subjects' L1). Grainger and Frenck-Mestre found a null effect of prime-target relationship in their lexical decision task, but when the same stimuli were used in a semantic categorization task, a significant priming effect was produced, as would be predicted by the Sense Model.

Finkbeiner et al.'s (2004) own research has also provided several key pieces of evidence that are consistent with their account. First, Finkbeiner et al. successfully replicated Grainger and Frenck-Mestre's (1998) results, finding a robust masked L2-L1 translation priming effect in semantic categorization, but not in lexical decision. These findings have also been replicated in more recent experiments (e.g., Wang & Forster, 2010; Xia & Andrews, 2015). Perhaps more compelling, however, is that Finkbeiner et al. tested the Sense Model in a within-language setting by pairing many-sense words (e.g., head) with semantically similar few-sense words (e.g., skull), and used both a many-to-few priming direction (i.e., head-SKULL) and a few-to-many priming direction (i.e., skull-HEAD), in both a lexical decision task and a semantic categorization task. Finkbeiner et al. found that, even in a within-language task, a significant priming effect was obtained in the many-to-few direction, but no priming was obtained in the few-to-many direction in lexical decision. In semantic categorization, on the other hand, priming was obtained in both directions, consistent with the Sense Model's predictions.

Despite the Sense Model's ability to account for these findings, there are several empirical challenges to its viability. Xia and Andrews (2015), for example, compared priming effects in the L1-L2 and the L2-L1 direction using both lexical decision and semantic categorization. While Xia and Andrews found that the priming effect was larger in semantic categorization than it was in lexical decision, replicating previous findings (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010), they also found that there was still a priming asymmetry in semantic categorization. Priming effects were still larger in the L1-L2 direction than in the L2-L1 direction, contrary to the assumptions of the Sense Model.

Another serious challenge for the Sense Model comes from Chen et al. (2014). Chen et al. conducted three lexical decision tasks, with the first two directly testing the predictions of the Sense Model in a bilingual setting. First, Chen et al. had Chinese-English bilingual subjects perform a lexical decision task, where the masked primes were polysemous English words, and the Chinese targets were single-sense words. Critically, these polysemous English words were defined based on the number of senses mastered by the subjects. Chen et al. had a group of subjects with similar English proficiency to their experimental subjects rate the number of senses of each English word. Words that had two or more senses based on these ratings were included as primes in their first experiment. Under the assumptions of the Sense Model, such primes should produce robust priming effects, as the primes should have activated all the senses associated with the targets. Second, Chen et al. had subjects perform a lexical decision task using single-sense L1 primes and polysemous L2 targets. Again, the Sense Model is clear in its predictions: L1 primes should not produce a robust priming effect if the proportion of primed to unprimed senses is low, which was the case in this second experiment.

Neither of these predictions were supported by Chen et al.'s results. First, even when using polysemous L2 primes and single-sense L1 targets, the priming effects were still null. Second, even under circumstances where the L1 prime would only prime a small proportion of the L2 senses, the priming effect still emerged. In short, even under conditions when the priming asymmetry should not occur, or, if it did, it should have been a reverse asymmetry, the same priming asymmetry was still observed.

Chen et al. then proposed an alternative explanation, arguing that, rather than being due to asymmetries between L1 and L2 words at the semantic level, the null priming effects are a result of the language dominance. In their experiments, Chinese was the native language of subjects, and there was a processing advantage compared to English. As such, the semantics of the L1 primes can be accessed faster than for L2 primes. To produce priming effects in the non-dominant language, then, more processing time would need to be devoted to an L2 prime. To test this prediction, Chen et al. conducted a final experiment in which English primes were presented for 250 ms, to guarantee that subjects would have enough time to access the semantics of the L2 prime. Their final study produced a sizeable (33 ms) translation priming effect.

Note further that the Sense Model also fails to take the proficiency of bilinguals into account. Whereas some accounts of bilingual language processing assume that proficiency affects the ease of access to conceptual representations from lexical-level representations, and predict that more proficient bilinguals should produce masked translation priming effects in the L2-L1 direction (e.g., Dijkstra & van Heuven, 2002; Kroll & Stewart, 1994), the Sense Model is not able to accommodate such a prediction. Instead, the Sense Model would predict the opposite: as bilinguals become more proficient in their L2, the senses that are acquired will tend to be language-specific. As a result, not only should L2-L1 priming still not occur, but L1-L2 priming should be reduced as well, as there would be less sense overlap in the semantic representations of L2 words for proficient bilinguals than for less proficient bilinguals, and a lower proportion of the senses in such words should be preactivated by L1 primes. Overall, while able to offer a very straightforward and understandable explanation of several findings in the literature, recent research has demonstrated serious flaws in the Sense Model. How these issues have been dealt with will be discussed after a review of some of the other theoretical accounts below.

1.3.4 The Revised Hierarchical Model

Perhaps one of the most cited models in all of bilingualism research, the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll & Tokowicz, 2001) was designed as a general-purpose model of bilingual memory rather than as an account of the masked translation priming asymmetry. The RHM assumes that words in a bilingual's two languages are stored in separate lexical memory systems, but share a common conceptual memory system. The two languages are also assumed to have bidirectional inter-lexical connections to each other, and access to each language is selective, such that bilinguals can inhibit or activate one language depending on the context. While words in either language can access conceptual representations, the RHM assumes that this ability differs for L1 and L2 words, depending on the strengths of the links between lexical and conceptual representations. For L1 words, conceptual representations can be readily accessed directly from the lexical forms, as it is assumed that the links between concepts and L1 word forms are very strong. For the L2, however, it is assumed that the direct conceptual links are weaker. There is thus an asymmetry in the connection between each lexicon and the conceptual representations. As a result, accessing meaning from L2 words often requires mediation by the L1 lexical representations. Thus, the lexical links from L2 to L1 are assumed to be much stronger than from L1 to L2, as the L2 is assumed to rely more on L1 for conceptual mediation than L1 does on L2. Over time, as bilinguals become more proficient in their second language, direct conceptual links are also acquired, and strengthen with L2 practice. Thus, this model assumes that, as bilinguals gain greater proficiency in their L2, their ability to directly access conceptual representations from their L2 increases.

The RHM (Kroll & Stewart, 1994) accounts for the translation priming asymmetry by assuming that the locus of the translation priming effect is at the conceptual level, rather than the lexical level, and since L2 lexical forms have a weaker connection to these representations, these primes do not effectively activate their conceptual representations, which means that the conceptual representations of the L1 targets are often not preactivated enough by an L2 prime. As a result, there are no priming effects. On the other hand, because L1 words have strong connections between their lexical forms and conceptual representations, L1 primes are effective at preactivating the conceptual representations of L2 targets. In addition, this account predicts that as bilinguals become more fluent in their L2, priming effects should begin to emerge in the L2-L1 direction, as L2 words should be able to preactivate the conceptual representations of L1 targets.

1.3.4.1 Empirical support for the RHM

Several findings have been interpreted as evidence for the RHM. Perhaps the most compelling evidence for the RHM comes from research done on balanced bilinguals. Up until this point, all the research that has been discussed has focused on bilinguals who acquired their languages at different periods in time. However, research on bilingual language processing has also been carried out on bilinguals that learned their two languages simultaneously from an early age. Unlike unbalanced bilinguals, balanced bilinguals are essentially equally proficient in their two languages. According to the RHM, the translation priming effect size should be comparable in the L1-L2 and the L2-L1 directions for balanced bilinguals. This prediction has been directly tested by Duñabeitia, Perea, and Carreiras (2010). Duñabeitia et al. tested highly fluent Basque-Spanish balanced bilinguals in both the Basque-to-Spanish and the Spanish-to-Basque direction using both cognates and noncognates. In addition to replicating the cognate priming advantage found in prior studies (e.g., Gollan et al., 1997; Sanchez-Casas et al., 1992, Experiment 1), Duñabeitia found that, unlike unbalanced bilinguals, balanced bilinguals do not show asymmetric priming effects. These results provide support for the RHM’s predictions that balanced bilinguals should produce symmetric priming effects, as lexical forms from both languages should be able to access conceptual representations with nearly equal efficiency.

Similar results to Duñabeitia et al.'s (2010) had previously been reported in interlingual semantic priming tasks where the primes and targets were not direct translation equivalents. Perea, Duñabeitia, and Carreiras (2008), for example, tested highly fluent Basque-Spanish balanced bilinguals in both the Basque-to-Spanish and the Spanish-to-Basque direction using associatively-related noncognate pairs, rather than translation equivalents. Using this design, Perea found a significant semantic priming effect for both Basque-to-Spanish and Spanish-to-Basque pairs. Contrary to the results obtained by de Groot and Nas (1991) with unbalanced bilinguals, the semantic priming effect was similar in size in the two directions.

The results reported by Chen et al. (2014) can also be explained by the RHM. When using masked primes, the asymmetry can be explained by the RHM's assumption that connections between L2 lexical forms and conceptual representations are weaker than the conceptual connections for L1 lexical forms. In their Experiment 3, the fact that priming effects emerged in the L2-L1 direction can be explained within the RHM framework by simply assuming that more time is needed to activate semantic representations from L2 lexical representations. Thus, the overall pattern of results reported by Chen et al. can be explained as being due to how easily the lexical forms in L1 and L2 can access their conceptual representations.

The assumption that priming effects should emerge in the L2-L1 direction as L2 learners develop greater proficiency in their L2 has also been directly tested in several empirical studies. The first investigation of the effects of proficiency on L2-L1 priming effects in unbalanced bilinguals was reported by Dimitropoulou et al. (2011a), who tested three groups of unbalanced Greek-English bilinguals, who had different L2 proficiency based on both subjective and objective measures of proficiency. What was unusual about this study was that there were priming effects for all three groups, and L2 proficiency did not modulate the size of the priming effect. Such results are, understandably, not consistent with any of the prior literature, nor were these results consistent with any account of bilingual word recognition. However, in a subsequent paper, Nakayama et al. (2016) noted that Dimitropoulou et al.'s measure of proficiency, the Cambridge ESOL, was problematic.

The issue is that the Cambridge ESOL allows an overlap in proficiency across its proficiency categories. Bilinguals can take the low-, intermediate-, or high-proficiency Cambridge ESOL tests, and proficiency is indexed by their performance on the test that they took. Under this testing system, a bilingual who struggles, but passes, the high-proficiency category test would still be rated as being more proficient than a bilingual who easily passed a lower-proficiency category test, but never took the high-proficiency category test. Instead of using the Cambridge ESOL, Nakayama used the TOEIC. The TOEIC is a standardized test of English proficiency that assesses English listening, reading, speaking, and writing skills for workplace environments, and was designed to better differentiate between L2 proficiency groups. Using the TOEIC, Nakayama et al. conducted lexical decision tasks, and found significant priming effects with highly proficient Japanese-English bilinguals, but found null priming effects with less proficient bilinguals. The RHM can effectively account for these findings if it is assumed that proficiency modulates the strength of the connections between L2 and the conceptual store. With greater proficiency, the access of conceptual representations by L2 lexical forms becomes more efficient. Hence, a priming effect is observed for highly proficient bilinguals.

1.3.4.2 Empirical Challenges to the RHM

Despite the considerable support for the RHM, the model is not without its empirical challenges. In particular, a review of the RHM by Brysbaert and Duyck (2010) discussed several findings which they argued present enough of a challenge to the RHM, in particular its assumption concerning selective access to the desired lexicon, to warrant abandoning the model in favour of the Bilingual Interactive Activation plus model (BIA+; Dijkstra & van Heuven, 2002; see below). Brysbaert and Duyck questioned several of the assumptions of the RHM, including the assumption that languages reside within separate lexical systems.

As evidence against the assumption of separate lexical systems, Brysbaert and Duyck cited Spivey and Marian's (1999) results. Spivey and Marian evaluated whether Russian-English bilinguals would be influenced by their knowledge of English while carrying out instructions based on auditory L1 words. This study used a visual world paradigm, in which subjects simultaneously view a few objects (e.g., a candy, an apple, a candle, and a fork) and are asked to assume that they were performing an action on one of the objects in response to a request to do so (e.g., "pick up the candle"). Spivey and Marian then tracked the eye movements of subjects to see what objects the subjects fixated on. When done in English, subjects often looked at the candy before the candle, consistent with Marslen-Wilson's (1987) cohort model of auditory recognition, which assumes that words starting with the same sounds are simultaneously activated, and only once more information is available are alternative, incorrect words eliminated.

Using this paradigm, Spivey and Marian (1999) gave Russian-English bilinguals instructions in L1 such as "Положи марку ниже крестика/*Poloji marku nije krestika*", or in their L2 "Put the stamp below the cross". One of the distracter items would be, for example, a marker (called a фломастер/*flomaster* in Russian). For Russian-English bilinguals, the words for marker and stamp would be competitors of each other, as the word for stamp (*marku*) sounds like the English word *marker*. Spivey and Marian found that subjects would often look at the marker before picking up the stamp. Overall, these results suggest that the names of objects in a bilingual's other language are activated even in monolingual experimental settings. Spivey and Marian's findings have subsequently been replicated several times (e.g., Marian, Blumenfeld, & Boukrina, 2008; Marian & Spivey, 2003; Marian, Spivey, & Hirsch, 2003).

The RHM's assumption that language access is selective has also been challenged by Brysbaert and Duyck (2010), who cited Dijkstra, Timmermans, and Schriefers's (2000) results. Dijkstra et al. adopted a go/no-go paradigm for use with Dutch-English bilinguals. In this experiment, their subjects were presented with words in English and Dutch, and subjects had to respond with a button press if an English word appeared, but had to wait for the next word if the word was Dutch. Dijkstra et al. compared words that existed only in English (e.g., *home*) to words that were interlingual homographs – words that exist in both languages, but have different meanings in the two languages (e.g., *room* means *cream* in Dutch). If subjects were able to selectively access their English lexicon while inhibiting their Dutch lexicon, subjects should not be influenced by whether the target had a meaning in both languages. Dijkstra et al. found that, regardless of whether subjects were tested in L1 or L2, subjects responded more slowly to interlingual homographs than non-homographs.

Other research has shown that, while lexical access appears to be nonselective in general, the nonselectivity of lexical access can be constrained by a number of factors. For example, Libben and Titone (2009) studied the effects of sentence constraint, defined as the extent to which the sentence context preceding the target word biased that word. French-English bilingual subjects read English sentences which either contained cognates (e.g., *piano*), interlingual homographs (e.g., *coin*), or matched control words, and the sentences either provided a low or high semantic constraint on the target language meaning. Under low semantic constraints, a significant cognate facilitation effect was found for first fixation, gaze duration, skipping, go-past time, and total reading times, while interlingual homographs produced inhibition. Under high semantic constraints, only early-stage measures (i.e., first fixation duration, gaze duration, and skipping) of comprehension were affected, suggesting that nonselective access is limited to early stages of comprehension in highly constrained contexts. Such results were consistent with other studies that have shown that contextual constraints place limits on nonselective lexical access (e.g., Duyck, Van Assche, Dreighe, & Hartsuiker, 2007; Schwartz & Kroll, 2006; van Hell & de Groot, 2008).

Kroll, van Hell, Tokowicz, and Green (2010) have more recently addressed some of Brysbaert and Duyck's (2010) criticisms of the RHM. While acknowledging that the RHM did originally assume selectivity, Kroll et al. noted that Kroll and de Groot (1997) discussed how the RHM could accommodate evidence for nonselectivity, and also noted that language selectivity was not a central issue that the model was created to address. Further, Kroll et al. note that such a critique of the RHM does not acknowledge that parallel access does not necessarily imply an integrated lexicon.

Regardless of whether Brysbaert and Duyck's (2010) critique of the RHM's assumptions of separate lexicons and nonselective lexical access carry any theoretical weight or not, the issue with the RHM that is most relevant to the current discussion is how the RHM can account for task-specific effects on L2-L1 translation priming. Given that studies typically find significant L2-L1 priming effects in semantic categorization tasks (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), there is no reason for the RHM to predict that the same subjects should not produce priming effects in another task such as the lexical decision task. That is, as Finkbeiner et al. argued, if the weak L2 form-meaning connections are not a limiting factor in one task, then they should not be a limiting factor in another task. The RHM thus has difficulty accounting for the task-specific nature of the priming asymmetry effect, and would require some modifications to successfully account for such findings.

1.3.5 The BIA+ Model

As with the RHM, the BIA+ (Dijkstra & van Heuven, 2002) arose as a general model of bilingual language processing. The BIA+ assumes that word processing in psychological experiments involves two subsystems: a word identification subsystem, and a task/decision subsystem. The word identification subsystem is comprised of units representing sublexical and lexical orthography and phonology, as well as semantics, and nodes denoting language membership. During the process of reading, nodes representing the sublexical orthography of words are initially activated, and contain bidirectional connections with both lexical orthography and sublexical phonology, both of which share their own bidirectional connections with lexical phonological units. Both lexical orthography and phonology, in turn, activate the semantic representations of the words and the language nodes. This information is then used by the task/decision subsystem, which determines the actions required to perform for the task.

Dijkstra and van Heuven (2002) make several assumptions regarding the word identification subsystem. First, contrary to the RHM, the word identification subsystem is assumed to have an integrated lexicon. Access to word representations in both languages is parallel and nonselective, in that words in both languages are activated when bilinguals are exposed to a stimulus. As a result, written words in one language can activate the orthographic, phonological, and semantic representations of the other language also, especially when the two languages share a common orthographic system. For example, for a French-English bilingual, the English word *four* can not only activate its translation equivalent in French, *quatre*, but also its interlingual homograph in French, *four*, which means *oven*, as well as any other similarly pronounced or spelled words in both English and French. Second, the word identification subsystem additionally has language nodes which denote the language membership of words based on information from lexical orthography and phonology. While these nodes are assumed to have no effect on the actual activation levels of word representations, the nodes are assumed to minimize the amount of interference from the nontarget language when bilinguals are processing in one of their languages. Finally, it is assumed that representations in the word identification subsystem differ in terms of their resting-level activations. Because bilinguals are typically more proficient in their L1 than their L2, representations for L1 are assumed to have higher resting-level activations than L2 representations. As a result, L1 representations require less time to become activated than L2 representations. However, as with the RHM, the BIA+ model assumes that the resting-level activations of L2 representations increase as a function of the frequency of use of the L2, and the bilingual's proficiency in the language.

1.3.5.1 Empirical Evidence for the BIA+ Model

Some of the earliest evidence consistent with the BIA+ model comes from research on orthographic neighborhood (Coltheart, Davelaar, Jonasson & Besner, 1977) effects on bilingual word recognition. Such results, in fact, provided some of the earliest evidence for the BIA+ model's predecessor, the Bilingual Interactive Activation model (BIA; Dijkstra & van Heuven, 1998). Using Dutch-English bilinguals, van Heuven, Dijkstra, and Grainger (1998) conducted both progressive demasking and lexical decision experiments to study how the recognition of words that belong exclusively to one language is affected by the word having orthographic neighbours (i.e., words that are spelled identically except for a single letter, meaning that log, fog, dig, dot, etc., are neighbours of dog) in either the same or the other language. Their results showed that responses to English targets were slowed by having a large number of orthographic neighbours in Dutch. When the number of neighbours was manipulated in the target word's language, inhibitory effects were consistently produced in Dutch, and facilitory effects were produced in English. These findings were interpreted as evidence that activation of word representations occurs in parallel in an integrated lexicon.

While making different assumptions about the organization of bilingual lexical memory, the BIA+ is often seen as being complimentary to the RHM, as the two models make similar predictions about masked translation priming effects. Much of the evidence discussed in the previous section on the RHM can also be said to be consistent with the assumptions of the BIA+ model. The BIA+ model can account for findings from studies on balanced bilinguals (e.g., Duñabeitia et al., 2010; Perea et al., 2008) if it is assumed that the resting-level activations of representations in bilinguals' two languages are similar. When the resting-level activations of the two languages are similar, there is no delay in the activation of L2 representations compared to

L1 representations. Thus, both languages can successfully activate the representations of their translation equivalent targets when used as masked primes, and no priming asymmetry should be observed. Further, evidence consistent with the assumption that the activation of the L2 is slower was seen in Chen et al.’s (2014) Experiment 3, where it was found that priming effects emerged in the L2-L1 direction, but only when the presentation time of the prime was increased. That is, from the perspective of the BIA+ model, such a result is accounted for by assuming that because L2 representations have lower resting-level activity they take longer to sufficiently activate. Nonetheless, L2 primes are able to activate the lexical and semantic representations of the L1 translation equivalent when given enough time, resulting in a significant priming effect. It is for this same reason that the BIA+ model is also well equipped to account for the effects of proficiency on masked priming.

The results of Nakayama et al. (2016), clearly showing L2-L1 priming for highly proficient bilinguals, can be easily accounted for by this model if it is assumed that proficiency increases the resting-level activity of L2 representations, increasing the efficiency with which these words are able to activate the representations of the L1 target. In addition, the BIA+ model can account for findings that present a challenge to the RHM. For example, much of the research by Marian and colleagues (e.g., Marian et al., 2003, 2008; Marian & Spivey, 2003; Spivey & Marian, 1999) that found evidence that the lexicons of bilinguals’ L1 and L2 are integrated is accounted for by the BIA+ model’s integrated lexicon assumption. Further, the BIA+ model can account for Dijkstra et al.’s (2000) results showing evidence of nonselective access of languages during monolingual tasks. The BIA+ model’s ability to account for such findings when those findings have been argued to present a challenge for the assumptions of the RHM have led some researchers (e.g., Brysbaert & Duyck, 2010) to argue that the BIA+ model should be the dominant model of bilingual word recognition.

1.3.5.2 Empirical Challenges for the BIA+ Model

Although the BIA+ model can provide a coherent account of the priming asymmetry effect in lexical decision, it remains unclear how the BIA+ model would account for the significant L2-L1 translation priming effects in both the semantic categorization task (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015) and the speeded episodic recognition task (e.g., Jiang & Forster, 2001) in situations where no priming is found in lexical decision. As with the RHM, the BIA+ model does not have an apparent mechanism to account for task-specific differences in translation priming effects. Although the model does have a task/decision system, it is assumed that task context cannot exert a top-down influence on processes in the word identification subsystem, as the actions executed by the task/decision system are based on the activation information from the word identification subsystem in a bottom-up manner. If priming cannot be observed for a set of bilinguals in the lexical decision task because the L2 activation was too slow or too weak to sufficiently activate the representations of the L1 translation equivalent, then there should also be no priming in other tasks such as the semantic categorization task or the speeded episodic recognition task.

What is clear about the RHM and the BIA+ model is that while both models make assumptions about the nature of bilingual language processing that are well supported by empirical studies, neither model can provide an adequate account of the flexible nature of task-specific priming effects observed in prior literature. As with the RHM, to account for these findings, the BIA+ model would require some modifications to allow processing to be influenced by the nature of the task context.

1.4 The Present Research

As the above discussion indicates, much of the research that has been reported (e.g., Finkbeiner et al., 2004) has assumed that decisions in both the lexical decision and semantic categorization tasks are based on activity at the semantic level of processing. However, such an assumption may be inappropriate, and other theorists have proposed that tasks differ with respect to the locus of processing where decisions are made. In monolingual research, an example of such an account was proposed by Balota, Ferraro, and Connor (1991). Balota et al.'s account assumes that there are distinct sets of reciprocally connected units that process phonological, orthographic, and semantic information. Critically, Balota et al. assumed that decisions in different tasks are based on the processing of different sets of units. For the lexical decision task, the locus of decision-making is based on activity within the orthographic layer. For tasks such as naming, the locus of the decision is in the phonological units. Finally, the semantic units are assumed to be the locus of semantic categorizations. Critically, it is assumed that any semantic influence on processing in tasks such as lexical decision and naming occur via feedback from semantic units to either orthographic units or phonological units, which is assumed to enhance the settling of units in these layers (e.g., Hino & Lupker, 1996; Pexman & Lupker, 1999). Such an account stands somewhat in contrast to the assumptions of the BIA+ model, which assumes that decisions are based on a task/decision subsystem, and that decisions are not based on activity in any specific layer of units.

If a major difference between tasks is the nature of the representation used to complete them, a claim that one could offer is, in the case of bilingual versions of these tasks, the influence of L2 on task performance is mainly related to L2 competency in domains that are critical to performing the task and, hence, that is the reason for the task differences. For example, in lexical decision, which is heavily based on processing at the orthographic level, perhaps it is the subject's knowledge of L2 orthography that predicts L2-L1 priming, rather than just overall proficiency. If such were the case, one might expect that L2-L1 priming effects in lexical decision would be predicted by subjects' receptive and expressive abilities in their L2. On the other hand, in a semantic categorization task, in which the semantic layer is the locus of the decision, while one might still expect that, although reading and writing abilities are important in predicting priming effects, perhaps priming effects are predicted more by subjects' semantic knowledge. While semantic knowledge may be difficult to quantify, one may look at subjects' patterns of L2 usage. For example, subjects who use their L2 more of the time in home, school, and other settings may have more opportunities to gain a richer representation of the meaning of L2 words. As such, one might expect the influence of L2 in semantic categorization to be predicted by factors such as the amount of time that the L2 is used in daily life and different social settings, as well as the speaking proficiency of the learner.

One of the shortcomings of prior research (e.g., Nakayama et al., 2016) was that proficiency was scored as a unidimensional construct. While there is no doubt that the TOEIC is a valid measure of English proficiency, using the total TOEIC score as a measure of (L2 English) proficiency is not optimal. Specifically, the TOEIC may gloss over differences in proficiency that subjects may have across different domains of English language use. For example, learners may be strong at speaking, listening, and reading in English, but their writing abilities may be weak. While such a learner's TOEIC score would likely be lower than a learner who is proficient across all these domains of English competency, the use of the total TOEIC score would not provide information on what domain of English proficiency may be weaker than the others. Note also that, for as much as these measures of English proficiency can be informative about learners' competency in the English language, they are not informative about the social contexts in which learners are using their L2, which may shape how the language is acquired. As it is possible that different tasks emphasize the use of different language skills, priming effects in different tasks may be dependent on different facets of L2 competency.

Such an expectation is not without foundation. Even in monolingual studies, research has consistently shown task differences in both neuroimaging and behavioural data. In neuroimaging research, for example, it has recently been shown by Chen, Davis, Pulvermüller, and Hauk (2013) that performing semantic categorizations is associated with greater activity in the left inferior frontal gyrus than performing lexical decisions, while performing lexical decisions is associated with greater activity in the right precentral gyrus, and reduced activity in the bilateral posterior middle temporal lobe. In behavioural data, a common finding is that different factors produce different effect sizes in different tasks. One of the most notable and often cited examples is the word frequency effect. The effect of word frequency is usually found to be one of the most robust predictors of lexical decision performance of any factors examined (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Brysbaert, Stevens, Mandera, & Keuleers, 2016; Brysbaert et al., 2011; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Monaghan, Chang, Welbourne, & Brysbaert, 2017; Spieler & Balota, 1997; Yap & Balota, 2009). Yet, in tasks such as naming and semantic categorization, research has shown that that the effects of frequency are somewhat small (e.g., Balota & Chumbley, 1984). Such results are often interpreted as evidence that decision processes in different experimental paradigms emphasize the use of different kinds of information to complete the task, even when the same manipulation is being used, for example, masked semantic priming (e.g., de Wit & Kinoshita, 2014a, 2014b, 2015). Extending this notion to factors which affect the ease of access to the lexical or semantic representations of primes, then, implies that it would not be surprising if the factors and language processing skills required to effectively use the prime to drive decisions on the target also differed across tasks.

In a task such as lexical decision, where subjects need to differentiate between words and nonwords, one factor that may affect translation priming performance is a knowledge of the nuances of the English language, and being able to use English to communicate and articulate ideas in a precise manner. This notion is very similar to Swain's (1995, 2000) Output Hypothesis. Swain argued that producing language, whether in the written or spoken modality, forces learners to process a language more deeply than required for inputting language, because it requires actively constructing the forms and meanings of the language. Undertaking a production task then, causes learners to notice gaps in their ability to express the precise meanings of things they wish to communicate. As a result, when trying to produce language, speakers/writers learn how to fill in gaps in their knowledge.

Given the orthographic nature of the lexical decision task, one would thus expect performance in such a task to be related to the knowledge of orthographic forms in one's L2, and in their L2 vocabulary. One skill which has been linked to L2 vocabulary knowledge is L2 writing competency (e.g., Coxhead, 2011, 2018; Johnson, Acevedo, & Mercado, 2016; Laufer & Nation, 1995; Staehr, 2008; Zhong, 2016). To express oneself competently in one's L2 in writing, a writer must not only know what words can be used in a sentence, but also how to use these words appropriately. Thus, understanding the range and constraints of word meanings leads not only to stronger productive abilities in text within a language, but also an enriched representation of certain aspects of the language in memory (see also Perfetti, 2007).

In a task such as semantic categorization, on the other hand, the task is typically characterized as being one that emphasizes semantic coding to a great extent (e.g., Hino, Lupker, & Pexman, 2002). It may be the case that, beyond any self-reported reading, writing, speaking, or listening skills in English, the acquisition of greater semantic knowledge is associated with the extent of usage of the language in everyday life, as one acquires greater knowledge of the meaning of words through real-world interactions with not only other individuals, but also with the objects and concepts associated with their L2 labels, creating a more enriched and crystallized understanding of what these labels mean.

Beyond accounting for how subject-specific differences in L2 proficiency contribute to translation priming, another issue that must be considered is the item-specific factors that contribute to the ease of access to the prime in masked translation priming tasks. One of the most testable predictions of the BIA+ model (Dijkstra & van Heuven, 2002), for example, is that the temporal delay of L2 activation is related to the resting-level activation of the word representations in the language. Such resting-level activity is affected by not only the learner's knowledge of their L2, but also by the characteristics of the words themselves. Some words are used more frequently in an L2 than others, and, as a result, would have a higher resting-level activation than lower-frequency words. It follows that such word-level differences would have an impact on the prime's ability to preactivate the target's representation. The frequency of the targets used in an experiment is another factor that would affect the size of priming effects produced, with recent research showing evidence that priming effects are larger when low-frequency targets are used than when high-frequency targets are used, and when bilinguals are less proficient in the target language than when they are more proficient in the target language (e.g., Nakayama, Sears, Hino, & Lupker, 2012, 2013; see also Nakayama et al., 2016). Such results are consistent with the idea that the facilitation that is associated with translation priming is larger when the processing of targets is more difficult. What is unknown from prior research, however, is whether item-based factors play the same role in mediating translation priming in different tasks, such as semantic categorization and episodic recognition, given that research has shown that the decision processes associated with different tasks emphasize the use of different types of information (e.g., Balota & Chumbley, 1984; de Wit & Kinoshita, 2014a, 2014b, 2015).

While much of what has been discussed in this review has focused on the translation priming asymmetry, the primary focus of the present research is instead on why the translation priming effect differs as a function of the task. It is clear that there is a quantitative difference in the priming effects that one obtains in a semantic categorization task versus a lexical decision task. The question becomes whether this quantitative difference is the result of a qualitative difference in the factors that predict L2-L1 priming in each task. Understanding what processes drive the L2-L1 translation priming effect in each of these tasks could provide valuable insights into why these overall patterns of effects emerge in the extant literature.

The present research addressed these types of issues by examining the impact of subject-based and item-based factors on masked translation priming effects in three tasks: a lexical decision task, a semantic categorization task, and a speeded episodic recognition task. Subject-based factors included English and Chinese proficiency and the use of English in daily living, while item-based factors included prime and target frequency, length, and stroke count. Because few, if any, studies have systematically examined the specific subject- and item-based predictors of translation priming effects, in this research I chose the factors that were the most obvious starting points for several reasons. First, for the subject-based predictors, these predictors would encompass a broad set of language skills in both L1 and L2, and represent skills that are measured by most standardized measures of language proficiency. While L2 skills were of primary interest in understanding what factors drive efficient access of meaning from masked L2 primes, the possibility that translation priming effects were affected by target language proficiency also needed to be considered. Second, it was desirable to gain an understanding of the patterns of language use of each subject, as the improvement in L2 skills is predictably related to the frequency of use of the L2 in daily life. Third, given the generally robust nature of word frequency effects in behavioural studies, tracking the frequency of both the prime and the target could provide insights into factors that drive translation priming on an item-by-item basis. Fourth, prime length and target stroke count were included to account for the orthographic complexity of the primes and targets. While it is likely that other factors affect translation priming on a subject- and item-level basis, these factors were not included to keep the study design more parsimonious, as it was deemed more important to gain an understanding of how these fundamental skills and characteristics contribute to driving performance before further work can be done to elaborate on other contributing factors.

The primary goal was to expand the current understanding of the underlying mechanisms that drive masked translation priming effects, how these processes differ across task contexts, and whether these different tasks also differ in the linguistic skills and item characteristics that affect L2-L1 priming in them. A second goal was to provide empirical information concerning how bilingual word recognition processes differ across different stages and facets of L2 development, and how the structure of bilingual memory changes as proficiency across different dimensions increases. Examining how proficiency changes the nature of L2-L1 priming effects should provide insight into how proficiency alters the structure and organization of bilingual memory over time as the L2 continues to develop.

While the issue of the nature of L2 representations has been directly addressed in models such as the RHM, it has not been developed to the same degree in the BIA+ model (e.g., van Hell, 2002; Jacquet & French, 2002; however, see Dijkstra, Haga, Bijsterveld, & Sprinkhuizen-Kuyper, 2012). Further, the notion that proficiency plays an important role in L2-L1 priming appears to run contrary to the core assumptions of the Sense Model, as well as the Episodic L2 Hypothesis, which both assume that the lack of L2-L1 priming is a consistent phenomenon across all proficiency levels for unbalanced bilinguals. However, at least in the case of the Episodic L2 Hypothesis, Forster (personal correspondence) has suggested that that model could be amended to include the assumption that, with greater proficiency, L2 representations migrate from episodic to lexical memory. As Forster further suggested, this idea has an interesting prediction. As bilinguals gain more proficiency in their L2, their priming effects in the speeded episodic recognition task should, in fact, diminish because many L2 representations would have "migrated" from episodic memory to lexical memory.

In addition to providing an overall framework for understanding task differences in translation priming, there were three general ideas concerning the three tasks in question (lexical decision, semantic categorization, speeded episodic recognition) that were investigated. First, if the reason one often obtains null priming effects in the L2-L1 direction in lexical decision is because prior research has not accounted for subjects' L2 orthographic knowledge and proficiency, subjects who report having high receptive and/or expressive competency in written English should produce a significant L2-L1 priming effect, while subjects who report having poor receptive and/or expressive abilities in written English should not produce a L2-L1 priming effect. In addition, the priming effects should be impacted by the relative frequency of the primes and targets. Priming effects should be larger for targets preceded by high-frequency primes than low-frequency primes, and should also be larger for low-frequency targets than high-frequency targets. These predictions were tested in Experiment 1.

Second, if the degree of priming obtained in a semantic categorization task is based on subjects' semantic knowledge, it should be found that habits and behaviours which would lead to greater acquisition of L2 semantic knowledge should lead to priming in semantic categorization. Specifically, the extent to which subjects use English across different social contexts, and, to a lesser extent, their expressive abilities in written and spoken English should be key factors. Whether prime and target frequency would mediate translation priming in semantic categorization in the same way that it would in lexical decision, however, is less clear. These predictions were examined in Experiments 2 and 3.

Finally, as stated above, if L2 representations start off represented in episodic memory, but then transition to lexical memory as speakers gain greater proficiency in their L2, then subjects who are less proficient in their L2 across all domains should produce a significant priming effect in the speeded episodic recognition task, whereas subjects who are highly proficient in their L2 across all domains should produce a smaller or null priming effect in the speeded episodic recognition task. Additionally, one should observe an effect of prime and target frequency. Because high-frequency L2 words are more likely to gain established representations in lexical memory, the priming effect in the speeded episodic recognition task should be more likely to occur with low-frequency primes than high-frequency primes. These predictions were tested in Experiment 4.

These research questions were addressed using Chinese-English bilinguals as the target population. Chinese-English bilinguals were used for two reasons. First, most of the research that has reported a translation priming asymmetry effect has been done with bilinguals whose languages use different scripts (e.g., Finkbeiner et al., 2004; Gollan et al., 1997). Different script bilinguals were of greater interest due to this fact, as many of the task-specific differences that have been reported have been obtained under this circumstance (however, see Grainger & Frenck-Mestre, 1998). There are several critical differences between English and Chinese orthography which may influence the amount of translation priming produced by these languages. Most obviously, English uses an alphabetic orthographic system, while Chinese uses a logographic system. One of the critical differences between alphabetic and logographic systems is the way in which semantics maps onto orthography. For alphabetic languages such as English, the relationship between form and meaning is highly opaque, in that the individual graphemes within the system do not carry meaning, and there is only a weak overlap between orthography and morphology. As Yan, Zhou, Shu, and Kliegl (2012) have argued, this opaque mapping between orthography and semantics can mean that information about word meaning only becomes available at a later stage of lexical processing, and would have to be mediated by phonology. For logographic systems such as Chinese, however, each orthographic unit contains morphosemantic information. The mapping between orthography and semantics in Chinese is arguably closer than the mapping between orthography and phonology. As a result, accessing the phonology of a word is not necessary when accessing semantics.

A second reason that Chinese-English (as opposed to, for example, Japanese-English or Hebrew-English) subjects were used was for convenience. Chinese-English bilingual students represented arguably the largest population of multilinguals that were available, which made acquiring a sample of subjects quicker and more efficient than if a different cross-script bilingual population had been used.

Chapter 2

2 Experiment 1

2.1 Method

2.1.1 Subjects

One-hundred-and-three undergraduate students (76 female, 27 male) at the University of Western Ontario participated in Experiment 1 for course credit. Of these participants, 97 were right-handed, three were left-handed, one was ambidextrous, and two failed to disclose their handedness. Subjects ranged in age from 18 to 34 years old (M = 19.29, SD = 1.69). Five subjects were excluded from the analyses due to not filling out their Language Experience Questionnaires (LEQs) properly (4.85% of the total data), leaving a total of 98 subjects. Of these 98 subjects, 78 subjects reported speaking Mandarin and English as their two languages, while one subject reported speaking Cantonese and English, but knew simplified Chinese script. Nineteen subjects reported being trilingual. Three subjects reported speaking Mandarin, English, and Japanese, and were familiar with Japanese Kana and Kanji in addition to English and simplified Chinese. Thirteen subjects reported speaking Mandarin, Cantonese, and English, and were thus familiar with both simplified and traditional Chinese, as well as English orthography. Two subjects reported speaking Mandarin, English, and occasionally French. Finally, one subject reported speaking Mandarin, English, and Korean, and was thus familiar with simplified Chinese script, English orthography, and Korean Hangul. All subjects had normal or corrected-to-normal vision.

2.1.2 Stimuli

Experiment 1 involved a set of 100 word and 100 nonword Chinese targets, which were paired with 200 English word primes. All words and nonwords were composed of two Chinese characters. For the nonwords, the combination of characters was such that, while each character could have been a word on its own, the combination of the two characters was not (e.g., 石虎, or "rocktiger"). Word targets were primed by either an English translation prime or an unrelated prime, resulting in 50 items per cell for each subject. The unrelated primes consisted of English words which were translation equivalents of other targets in the experiment, that is, the pairs were created by re-pairing the unrelated primes and targets (e.g., game-衬套, bush-游戏). Two lists of primes and targets were created to ensure that each target appeared in each prime condition across all lists. Words and nonwords were matched on stroke count. Mean ratings for stroke count and log frequency for all targets, as derived from the Chinese Lexicon Project (Tse et al., 2017), can be found in Table 1. Every target used in Experiment 1 can be found in Appendix B, which also shows the translation and control primes which were paired with it.

Table 1. *Means and Standard Deviations for Prime CELEX, Prime Length, Target Log-Transformed Google Frequency, and Target Stroke Count for Words, Experiments 1-4.*

	Experiment					
	LDT		SCT		sERT	
Factor	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prime CELEX	36.30	121.98	30.74	68.51	50.55	59.29
Prime Length	5.81	1.41	5.76	2.07	5.76	1.42
Target Google Frequency	5.84	0.55	5.45	0.41	5.78	0.36
Target Stroke Count	22.57	6.90	22.33	7.09	20.91	5.27

Note: LDT = Lexical Decision Task; SCT = Semantic Categorization Task; sERT = Speeded Episodic Recognition task.

2.1.3 Apparatus

Stimuli were presented on an LG Flatron W2242TQ-BF LCD monitor, which had a refresh rate of approximately 60 Hz. Recording of response latencies and accuracies was done using DMDX software (Forster & Forster, 2003), with responses being made by pressing keys on a keyboard.

2.1.4 Procedure

Subjects read a detailed letter of information about the study, and then provided their informed consent. Information about the subject’s background – including their age, the amount of time spent living in Canada, and their IELTS score – was then obtained. Subjects then completed a questionnaire to assess their self-reported level of proficiency, and the contexts in which they have used and acquired English. Subjects then sat in front of a computer. Subjects completed both the LDT for Experiment 1, and the SCT for Experiment 2 (the details of which will be presented subsequently). Half of the subjects completed the LDT first, and half completed the SCT first. Verbal instructions were either given in English if the experimenter was an English monolingual, or in Chinese if the experimenter was a native Chinese speaker. Letters of information, consent, and questionnaires were also conducted in English. The instructions for each experiment were exclusively written in simplified Chinese script.

For Experiment 1, subjects were instructed to decide whether each target was a Chinese word or nonword as quickly and as accurately as possible, pressing the right Shift key for a word, or the left Shift key for a nonword. Subjects received 6 practice trials before beginning the experiment. The experiment itself consisted of a single block of 200 trials, with each trial beginning with a forward mask (############) for 500 ms, followed by the prime for 50 ms, then a backward mask (&&&&&&&&&&&&) for 150 ms, and finally the target to which they had to respond. As a result, the SOA was 200 ms, replicating the SOA used by Finkbeiner et al. (2004). All masks and primes were presented in 14-point Courier New font, while the Chinese targets were presented in 14-point DengXian font.

2.1.5 Measures

2.1.5.1 Background Information Questionnaire

This questionnaire collected basic demographic information, including age, gender, whether the subject was born in Canada or came from abroad, as well as the number of years that the subject had been living in Canada.

2.1.5.2 The Language Experience Questionnaire (LEQ)

This questionnaire was largely based on the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), which is a self-assessment measure involving several variables. The LEQ measures language exposure across several domains. First, subjects would indicate their native country, native language, and their second language, and then indicate at what age they moved to Canada if Canada was not their native country. Afterwards, subjects would indicate the order in which they learned their languages, and order the languages they know from most proficient to least proficient. Subjects were then asked about their use of English and Chinese in different environments and social contexts. Subjects gave estimates for the percentage of time that they used English and Chinese at home, at school, and in other social settings, and then rated their language proficiency across four domains: speaking, understanding, reading, and writing. Subjects also rated how proficient they were in both English and Chinese on a 10-point scale, ranging from 1 (very little proficiency in the language) to 10 (highly proficient in the language). The questionnaire took approximately 5 minutes to complete, and consisted of 21 questions. The reliability of these measures was found to be good, as the self-rated speaking, understanding, reading, and writing measures were internally consistent in both English (Cronbach's $\alpha = .92$), and Chinese (Cronbach's $\alpha = .83$), while the use of English at home, school, and in other social settings was relatively poor (Cronbach's $\alpha = .52$). The mean values for the LEQ can be found in Table 2 for Experiments 1-3, and Table 3 for Experiment 4.

Table 2. *Mean Language Experience Questionnaire Responses, IELTS, sPIP, iPIP, and PIP scores for Subjects, Experiments 1-3.*

	Experiment					
	Experiment 1		Experiment 2		Experiment 3	
Factor	*M*	*SD*	*M*	*SD*	*M*	*SD*
PEH	9.10	13.69 (60)	9.10	13.69 (60)	6.45	7.61 (100)
PES	65.53	25.41 (90)	65.53	25.41 (90)	70.07	22.82 (50)
PEO	36.46	29.75 (100)	36.46	29.75 (100)	43.94	29.10 (85)
ER	7.17	2.14 (7)	7.17	2.14 (7)	7.63	1.14 (5)
EW	6.34	2.10 (7)	6.34	2.10 (7)	8.03	1.21 (5)
EL	7.30	2.28 (7)	7.30	2.28 (7)	8.03	1.21 (5)
ES	6.73	2.27 (7)	6.73	2.27 (7)	6.43	1.60 (6)
CR	9.25	1.71 (3)	9.25	1.71 (3)	9.42	0.80 (4)
CW	8.63	1.92 (5)	8.63	1.92 (5)	8.65	1.71 (6)
CL	9.46	1.51 (4)	9.46	1.51 (4)	9.52	0.50 (1)
CS	9.38	1.53 (3)	9.38	1.53 (3)	9.47	0.80 (3)
IELTS	6.02	2.14 (5)	6.02	2.14 (5)	6.52	1.82 (4)
sPIP	73.25	15.91 (111)	82.99	29.63 (144)	28.65	21.77 (101)
iPIP	180.99	332.50 (1290)	71.13	97.80 (964)	40.68	106.43 (972)
PIP	0.00	1.00 (9.45)	0.00	1.00 (7.58)	0.00	1.00 (7.05)

Note: PEH = Percentage of time English is used at home; PES = Percentage of time English is used at school; PEO = Percentage of time English is used in other social settings; ER = Self-rated English reading proficiency; EW = Self-rated English writing proficiency; EL = Self-rated English listening proficiency; ES = Self-rated English speaking proficiency; CR = Self-rated Chinese reading proficiency; CW = Self-rated Chinese writing proficiency; CL = Self-rated Chinese listening proficiency; CS = Self-rated Chinese speaking proficiency; IELTS = International English Language Testing System; PC = Prime CELEX frequency; GF = Log-transformed target Google frequency; L = Prime length; NS = Target stroke count; sPIP = Subject Proficiency Impact on Priming; iPIP = Item Proficiency Impact on Priming; PIP = Proficiency Impact on Priming. The data reported for Experiment 3 use Experiment 2 coefficients. Ranges are shown in parentheses.

Table 3. *Mean Language Experience Questionnaire Responses, IELTS, sPIP, iPIP, and PIP Scores for Subjects, Experiment 4.*

Factor	*M*	*SD*
PEH	19.64	29.87 (100)
PES	73.78	22.74 (80)
PEO	48.27	31.30 (98)
ER	7.73	1.16 (5)
EW	6.87	1.33 (6)
EL	7.90	1.14 (5)
ES	6.81	1.68 (6)
CR	9.27	0.95 (3)
CW	8.44	1.72 (3)
CL	9.47	0.75 (2)
CS	9.38	1.00 (3)
FL	9.33	4.66 (16)
YL	11.69	5.48 (18)
IELTS	6.43	1.92 (4)
sPIP	62.53	29.96 (157)
iPIP	-599.50	218.53 (1381)
PIP	0.00	1.00 (5.20)

Note: PEH = Percentage of time English is used at home; PES = Percentage of time English is used at school; PEO = Percentage of time English is used in other social settings; ER = Self-rated English reading proficiency; EW = Self-rated English writing proficiency; EL = Self-rated English listening proficiency; ES = Self-rated English speaking proficiency; CR = Self-rated Chinese reading proficiency; CW = Self-rated Chinese writing proficiency; CL = Self-rated Chinese listening proficiency; CS = Self-rated Chinese speaking proficiency; FL = Age at which subject first learned English; YL = Number of years subject has been learning English; IELTS = International English Language Testing System; PC = Prime CELEX frequency; GF = Log-transformed target Google frequency; L = Prime length; NS = Target stroke count; sPIP = Subject Proficiency Impact on Priming; iPIP = Item Proficiency Impact on Priming; PIP = Proficiency Impact on Priming. The data reported for Experiment 3 use Experiment 2 coefficients. Ranges are shown in parentheses.

2.1.5.3 International English Language Testing System (IELTS)

The IELTS is a standardized test of English language proficiency that tests the ability of test takers to listen, read, write, and speak in English. The test has four parts: A listening module, a reading module, a writing module, and a speaking module. The test takes approximately 2 hours and 44 minutes to complete. Test takers receive a score for each module, using a nine-point scale. Each point corresponds to a specific competence level in English, with a 1 corresponding to a non-user, and a 9 corresponding to an expert user. The IELTS is typically used when enrolling in an academic institution in English-speaking countries. Thus, any international students participating in any of the present studies had scores from the IELTS. The mean IELTS scores for subjects can be found in Table 2 for Experiments 1-3, and Table 3 for Experiment 4. In general, the IELTS was found to positively correlate with self-rated reading, $r(82) = .42$, $p < .0001$, writing, $r(82) = .39$, $p = .0002$, speaking, $r(82) = .45$, $p < .0001$, and listening proficiency in English, $r(82) = .49$, $p < .0001$, indicating that these self-assessed estimates of L2 proficiency had good construct validity.

2.1.5.4 Item-Specific Factors

Prime CELEX frequency (Baayen, Piepenbrock, & Gulikers, 1995) and length were derived using the N-Watch program (Davis, 2005), while target stroke count and frequency were derived from the Chinese Lexicon Project database (Tse et al., 2017).

2.1.5.5 Proficiency Impact on Priming (PIP)

Although it was not the intent to use the IELTS score alone to differentiate between highly proficient and less proficient subjects, it would not have been possible in any case because subjects' IELTS scores were highly homogeneous, as the data were found to be highly leptokurtic, having a kurtosis of 6.21 (*SE* = 0.53), which indicated that the values of the IELTS score tended to cluster around the center of the distribution. For example, the most common score on the IELTS was 6.5 and over 50% of subjects scored 6.5 or lower on the IELTS. It was thus impossible to evenly divide subjects into separable groups using the IELTS alone, as any splitting of the data at the median would either require including subjects who scored 6.5 to belong to separate proficiency levels, or for subjects with a score of 6.5 or lower to be categorized as low-proficiency subjects, and subjects with a score greater than 6.5 as highly-proficient. Such a split would result in quite uneven groups. And, of course, using the IELTS score alone also glosses over valuable information about the context of language usage from the LEQ. Information about the use of English and Chinese in different social contexts, subjects' self-rated proficiency, as well as the amount of time immersed in English- versus Chinese-speaking environments were factors that needed to be included. Therefore, a new, transformed score, based on the set of information collected was created as a more complete measure of how L2 proficiency affects priming effects.

That is, the Proficiency Impact on Priming (PIP) measure was designed specifically to understand what factors affect the access of lexical and semantic information associated with L2 primes and L1 targets, as measured by each subject's and each item's outcome variable, their mean priming effect. As was discussed, while standardized measures of L2 proficiency such as the TOEIC have been shown to predict L2-L1 priming in lexical decision (e.g., Nakayama et al., 2016), such a measure is highly broad, and it is unknown whether L2-L1 priming is affected more by specific domains of L2 competency (e.g., reading, speaking, writing, understanding), or by the general proficiency of the L2 learner. One approach to resolving this issue would be to derive a set of weights using linear modeling, and then using those weights to compute a composite measure that can be used to predict the effect size of the priming effect. Such a problem can be addressed with multiple regression, but using a standard multiple regression runs into the problem of overfitting the data, and not providing a reliable predictive measure that can generalize to new data. Further, the inclusion of too many factors in an analysis also increases the risk of overfitting the data. The objective of the present research was to derive a set of factors that can predict L2-L1 translation priming beyond the sample collected.

One method for resolving these issues is to regularize the linear regression models. Regularization is a technique in machine-learning in which the coefficient estimates of predictors are constrained to as small values as possible, which discourages the model from fitting on overly complex patterns in the data, and avoids the risk of overfitting. Another method for resolving the issue of overfitting is extracting the most relevant features for predicting L2-L1 priming, and excluding irrelevant factors. Preferably, regularizing these models while simultaneously extracting the most relevant features would enable the differentiation of how various dimensions of L2 competency affect semantic access in masked L2-L1 priming through using feature weights while, again, preventing the model from overfitting the data. Further, because one of the purposes of the present research was to test whether the relationship between different domains of L2 competency and masked translation priming changes across task contexts, this method would allow for direct comparisons of the skills and L2 use patterns that predict priming in each task, by comparing the features extracted and the feature weights used in different tasks. Finally, such a method would allow one to study the contribution of both subject-specific factors (e.g., L2 competency) and item-specific factors (e.g., prime frequency, target stroke count, target frequency) on L2-L1 translation priming. PIP was created with these objectives in mind.

To compute PIP, a series of three machine-learning models were used: a ridge regression model (Hoerl, 1962; Hoerl & Kennard, 1970; Tikhonov, 1943, 1963; Tikhonov & Arsenin, 1977), a lasso regression model (Tibshirani, 1996), and an elastic net regression model (Zou & Hastie, 2005). Each of these models are a regularized version of the standard linear regression, and offer the advantage of constraining the model's weights to reduce overfitting, and are robust when dealing with the problem of multicollinearity (e.g., Duzan & Shariff, 2015; Muhammad, Maria, & Muhammad, 2013; Oyeyemi, Ogunjobi, & Folorunsho, 2015). Each model used subject-specific factors such as self-reported L2 speaking, writing, reading, and listening proficiency or item-specific factors such as prime and target frequency as predictors, and the mean priming effect for each subject or item as the outcome variable. The fitting was done for priming data for each task separately, as it was predicted that priming effects would be affected by different dimensions of proficiency depending on the task context. The PIP score represents a composite score, and is composed of two subscales that can be combined to produce an overall PIP score. The first subscale, sPIP, is a predictive measure that uses subject-based factors, such as subject proficiency, in making predictions. The second subscale, iPIP[1], uses item-based factors, such as prime and target proficiency, to make predictions.

[1] Prime and target frequency are not aspects of proficiency, making iPIP somewhat of a misnomer. However, to reinforce the idea that it does represent a parallel to the sPIP concept in terms of trying to predict performance, the term iPIP will be used throughout.

2.1.5.5.1 Ridge Regression

The first machine learning model that was fit to the priming data was a ridge regression model. Also known as the Tikhonov regularization (e.g., Hoerl, 1962; Hoerl & Kennard, 1970; Tikhonov, 1943, 1963; Tikhonov & Arsenin, 1977), the ridge regression is a regularized version of a standard linear regression. The ridge regression works by introducing a regularization term to the linear model's cost function. The result of adding this regularization term is that the learning algorithm must fit the data while keeping the weights of the model as small as possible. The constraint on weights was controlled by α. With an α of 0, a ridge regression would be the same as a linear regression, while having a large α would result in the weights being close to zero. A full, detailed explanation of the logic of ridge regressions can be found in Appendix A.

2.1.5.5.2 Lasso Regression

The least absolute shrinkage and selection operator (LASSO) regression is yet another form of the regularized linear regression model. Where the ridge regression and the lasso regression differ is in terms of the type of cost function that the model adds. In a ridge regression, the regularization term is computed as the square root of the sum of the squares of the coefficients that are associated with each vector, which is also known as an ℓ_2 norm regularization. With a lasso regression, the regularization term is computed based on the sum of the coefficients of each vector, also known as an ℓ_1 norm regularization. Further, unlike ridge regressions, where each predictor is assigned a weight that is greater than zero, lasso regressions tend to eliminate the weights of the least important features, reducing them to zero. As such, lasso regressions perform feature selection and assign weights only to the most important predictors (Tibshirani, 1996). The lasso regression was trained and validated using the same method as the ridge regression described above. The results of this process will be discussed in greater detail below. A description of how the cost function is computed, and a description of each of its hyperparameters, and the specific values of each hyperparameter are found in Appendix A.

2.1.5.5.3 Elastic Net Regression

As with both the ridge and lasso regressions, elastic net regressions force the model to fit the data while keeping the weights as small as possible. What makes the elastic net different is that it is a hybrid between the ridge regression and the lasso regression, and uses a regularization term that includes both the ℓ_1 regularization term associated with the lasso regression and the ℓ_2 regularization term associated with the ridge regression. A full description of the elastic net regression, including descriptions of its cost function and hyperparameters can be found in Appendix A.

2.1.5.5.4 Computing PIP

The PIP score was created as a composite score based on two subcomponent scores: sPIP, and iPIP. Both components were created after the collection of the data. The sPIP component measured subject-specific factors that contributed to the production of a translation priming effect, and used subjects' responses on the LEQ and their IELTS as predictors, and subjects' mean priming effects as the dependent variable. The iPIP component measured item-specific factors that contributed to the production of translation priming, and included factors such as the CELEX frequency of the prime, the Google frequency of the Chinese target, the prime's letter length, and the number of strokes each target was comprised of. This computation was done by using a multistep method. First, the mean priming effects were obtained for each subject and for each item in the relevant behavioural task. After the mean priming effects were obtained, two datasets were created for each experiment. The first dataset contained the mean priming effects by subjects, and the subject-specific predictors, which included the subject's IELTS score, the percentage of time English was spoken by subjects in the home, at school, and other settings, as well as self-reported English and Chinese speaking, reading, writing, and listening abilities. Experiment 4 included two additional factors: the length of time that the subject has been learning English, and the age at which subjects first acquired English. To ensure that each model accounted for differences in performance that were due to differences in L1 skill, Chinese proficiency was included in the model to ensure that the model's predictions were not confounded by L1 abilities. The second data set contained the mean priming effects by item, as well as the item-specific predictors, which included the prime's CELEX frequency and length, and the target's Google frequency and stroke count. Experiments 2 and 3 (the semantic categorization tasks) had subsequent analyses which contained an additional predictor: semantic category typicality ratings for each prime. All fitting was done using only the positive trials (i.e., words, exemplars, and old items).

In computing the sPIP and iPIP scores, the predictors were first rescaled using the StandardScaler() function in the scikit-learn library (Pedregosa et al., 2011) in Python 3.5.1 (Python Software Foundation, 2015), and the priming effects were mean centred. After rescaling the predictors so the values were in the same numerical range, the priming and predictor data were then split into a training and testing set. The training set was used to fit the models to the priming data and tune the hyperparameters of the model, and consisted of approximately 80% of the entire dataset. The testing set was used to validate that the predictions of the model generalized to new data. Once each model was fit on the training data, its predictions were compared to actual priming effects and error rates, and both the mean squared error (*MSE*) and the root mean squared error (*RMSE*) were computed.

Before validating the models on the testing set, the models needed to be evaluated for whether they were overfitting the training data. To ensure that the models were not overfitting the data, and that they would be well-tuned to deal with newer data, the models were further regularized by performing a randomized search to find the optimum combination of hyperparameters. Hyperparameters are parameters whose values are set before the learning process begins, rather than being derived through training. Tuning the hyperparameters of a model provides the benefit of minimizing the cost function, while ensuring that the model is not overfitting the data. Rather than manually experimenting with different hyperparameters to determine which hyperparameters regularize the model best, a randomized search was performed through a specified subset of the hyperparameter space of the models to select the best combination of hyperparameters for each model (Géron, 2017). This randomized search was then evaluated using a k-fold cross-validation method, which involved dividing the training data into ten subsets, or *folds*, of data, and then subsequently training and evaluating each model ten times, picking a different fold for evaluation every time, and training using the other nine folds. The fit of each iteration was evaluated using the normalized mean squared error (*NMSE*), which provides an estimation of the overall deviations between the predicted and actual values.

Whereas the *MSE* can be computed as $MSE = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{X}_i\right)^2$, the *NMSE* is instead computed as $NMSE = 1 - \frac{\sum_{i=1}^{n}\left(y_i' - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i' - \overline{y'}\right)^2}$, where yi' is the predicted value of y, and where $\overline{y'} = \frac{1}{n}\sum_{i=1}^{n} y_i'$. The consequence of normalizing the *MSE* is that the scores will now range between -Inf and 1, with a value of 1 indicating the best possible score. A well-optimized model is expected to have an *NMSE* that is close to 1 (e.g., Liang, Hamada, Oba, & Ishii, 2018). The k-fold cross-validation, in this instance, produces a total of ten *NMSE* scores per randomized search. The randomized search was carried out for five thousand iterations per model. The set of hyperparameters which produced the best model for each of the three models were selected. Finally, the new models were validated on the testing set, and a set of coefficients was derived.

Once all three models were tuned, trained, and validated, a final k-fold cross-validation was performed on each model using the testing data set using five folds, and the performance of each model for each iteration was scored using the *NMSE* of the predictions. The *RMSE*s were then derived from this final cross validation, and the mean and standard deviation of the *RMSE* for each model was then computed. Using the mean and standard deviations of the *RMSE* for each model then allowed a comparison of how each model performed, which was then used to assign weights to each model's coefficients. The mean and standard deviations of the *RMSE* for each model for each experiment type can be found in Tables 4 and 5.

Table 4. *Mean and Standard Deviations for the Residual Mean Squared Errors for Each Model for Each Experiment, sPIP scores.*

	Model					
	Ridge		Lasso		Elastic Net	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	39.47	11.46	39.71	11.75	39.49	10.92
2	35.23	17.51	36.37	16.29	35.32	16.85
3	16.49	7.10	5.80	5.56	14.17	4.33
2 & 3	43.85	10.17	41.13	9.39	44.06	10.26
4	34.69	12.58	38.40	12.47	34.87	11.37

Table 5. *Mean and Standard Deviations for the Residual Mean Squared Errors for Each Model for Each Experiment, iPIP scores.*

	Model					
	Ridge		Lasso		Elastic Net	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	45.62	9.43	47.39	11.50	45.50	9.26
2	37.33	9.89	48.38	7.23	47.37	4.79
3	103.60	20.43	124.04	39.67	104.87	23.30
2 & 3	32.03	14.30	32.56	13.87	32.06	14.31
4	176.86	4.95	174.97	5.13	177.07	5.13

PIP was computed using an ensemble method. In machine-learning, ensemble methods aggregate the predictions of multiple models into a single, final prediction. Such a method is common in both machine-learning regressors and classifiers (e.g., Diettrich, 2000). The purpose of using such a method is that ensemble regressors can often perform better than any single regressor, by capitalizing on the strengths of each model, and compensating for each model's weaknesses. For the purpose of PIP, a simple averaging method was used, where the final coefficients used to compute sPIP and iPIP reflected the weighted average of the coefficients derived from the three models that were fit. The best-performing model's coefficients were weighted three times as much as those of the other two models. The coefficients were then aggregated, and the mean coefficients for each predictor were derived. Using these coefficients, the sPIP and iPIP subscores were then computed by aggregating the weighted sum of the predictor values from the ensemble measure for each subject and each item in the experiment. These values were then scaled by mean centering the scores, and dividing the scores by the standard deviation. A composite PIP score was then computed by adding the scaled sPIP and iPIP scores, and once again scaling the measure. The mean unscaled sPIP, iPIP, and the scaled PIP scores for Experiments 1-3 can be found in Table 2, while the same data can be found for Experiment 4 in Table 3.

2.2 Results

2.2.1 Data Trimming

Before trimming the data, five subjects were excluded from the analyses because they failed to provide responses on the Language Experience Questionnaire (4.85% of the total data), meaning that their PIP score could not be computed. Data trimming was done in three steps for both the LDT and SCT. First, if any items or subjects had an accuracy below 50% on either the LDT or SCT, they were immediately excluded from any analyses. One item was excluded from the LDT analysis, and two items were excluded from the SCT analysis (0.75% of the total usable data). Five subjects (5.06% of the total usable data) were also excluded from the analyses because they had accuracy scores below 50% on either the SCT or the LDT. Next, subjects' overall performance and item performance for every item type in both experiments were screened for multivariate outliers in speed-accuracy space using a Mahalanobis distance statistic and a *p*-value cut-off of .01 (Mahalanobis, 1936). This technique is similar to the screening technique used by Armstrong and Plaut (2016). Doing so eliminated nine subjects (9.11% of the usable data), five items in the LDT, and five items in the SCT (3% of the usable data). While this method eliminated 12% of the usable data, it helped minimize the risk of the results being driven by specific items or subjects. Finally, after this screening, trials with latencies that deviated by more than 3 standard deviations from each subject's mean RT for each experimental condition or were faster than 250 ms and slower than 2000 ms (1.75% of the total data; see Van Selst & Jolicouer, 1994), and errors were removed (3.97% of the total data), leaving approximately 72% of the total latency data (76% of the total usable data).[2]

2.2.2 PIP

The coefficients for Experiment 1 can be found in Table 6. As seen from the table, the largest subject-based predictors of priming effects according to this model were self-rated listening and writing abilities in English, and self-rated speaking and listening proficiency in Chinese. Negatively associated were self-reported reading and writing abilities in Chinese. Additionally, prime CELEX frequency was found to be the only item-based factor to have a facilitative effect on L2-L1 priming, while target Google frequency was found to have a negative relationship with priming.

Table 6. *PIP Coefficients for Experiment 1.*

	PIP Coefficient Values
sPIP	
CS	8.30
EL	5.43
CL	2.07
EW	1.73
CR	-3.14
CW	-5.51
iPIP	
PCEL	4.70
GF	-3.37

Note: CS = Self-rated Chinese speaking proficiency; EL = Self-rated English listening proficiency; CL = Self-rated Chinese listening proficiency; EW = Self-rated English writing proficiency; CR = Self-rated Chinese reading proficiency; CW = Self-rated Chinese writing proficiency; PCEL = Prime CELEX frequency; GF = Target Google frequency.

[2] Analyses with looser criteria are reported in footnotes if the results were qualitatively different. Subject data were only removed from the experiment that they produced an error rate exceeding 50%, the Mahalanobis distance criterion was loosened to a critical value of .001, the lower limit for RTs was lowered to 200 ms, and the upper limit was increased to 3000 ms. Participants and items were only excluded if they were extreme speed-accuracy outliers. Four participants were removed as multivariate outliers instead of nine, and eight items were removed instead of ten. Doing so retained 81% of the overall data, and 85% of the total usable data.

2.2.3 Reaction Time Analysis

The raw response times were submitted to a generalized linear mixed effects model using R's (R Core Team, 2017) lme4 package (Bates, Maechler, Bolker, & Walker, 2015), with subjects and items treated as random effects (Baayen, Davidson, & Bates, 2008). Three separate RT analyses were conducted. First, an analysis was conducted using prime type and sPIP as fixed effects. A second analysis was conducted using prime type and iPIP as fixed effects. Finally, prime type and the composite PIP score were analyzed. In all cases, sPIP was treated as a random slope on items, iPIP was treated as a random slope on subjects, and PIP was treated as a random slope on both subjects and items, unless otherwise mentioned. Due to recent concerns with transforming RTs to make the data abide by the assumption of normality required by standard linear mixed effects analyses (e.g., see Balota, Aschenbrenner, & Yap, 2013; Lo & Andrews, 2015), generalized linear mixed effects modeling was used because such models allow for the distributional assumptions to be determined by the researcher, allowing raw RTs to be submitted to the analysis without transformation. The RT data were analyzed using an Inverse Gaussian distribution.

The Bayes information criteria (BIC) from each model was compared to the BIC of alternative models to calculate the Bayes factor (BF) for each comparison. The Bayes factor allows for the testing of alternative hypotheses within the design against the null hypotheses (e.g., Rouder, Morey, Speckman, & Province, 2012). As an example, consider two models that attempt to model the effects of concreteness, prime type, and the prime's CELEX frequency on response times: a full model in which all of the additive effects and the interaction effects are retained, and an additive model in which only the additive effects are included. To determine which model is more consistent with the data, the Bayes factor can be calculated by comparing the BIC of these models to each other using the following formula: $BF = exp\left[\frac{BIC(M_2)-BIC(M_1)}{2}\right]$ (Wagenmakers, 2007). If the Bayes factor value from this comparison of the full model to the additive model was 3.53, for example, this value would indicate that the data were 3.53 times more likely to occur under the full model than under the additive model. However, if the Bayes factor value from this comparison is .01, this value would indicate that the data were 100 times more likely to occur under the assumptions of the additive model than the full model. This method can thus be useful in evaluating the amount of supporting evidence for each model.

A second method used to evaluate each model was the relative likelihood (θ) of each model. The relative likelihood is measured by comparing the Akaike information criteria (AIC; Akaike, 1973, 1974) of two models, using the following formula: $\theta = exp\left[\frac{AIC(M_2)-AIC(M_1)}{2}\right]$ (Burnham & Anderson, 2002, 2004). The result of this comparison is again directly interpretable, and indicates the likelihood that each model would minimize information loss compared to the other model. For example, finding a relative likelihood of 7.32 would indicate that the full model is 7.32 times more likely than the additive model to minimize information loss.

In some circumstances, however, the results of the Bayes Factor and the relative likelihood may be in contradiction to each other. Consider the situation in which the models account for the effects of prime type and concreteness on response times. The additive model, in such a circumstance, might be favoured over the interactive model in the Bayes Factor (e.g., 3.53), but the interactive model could be favoured over the additive model in the relative likelihood (e.g., 7.32). In such a circumstance, the additive model is more likely to account for the trends in the data, but it does so with a greater likelihood of data loss. Further, suppose that in the interactive model, the two-way interaction and the effect of prime is significant, but the effect of concreteness is not. In such circumstances, it is possible that the data loss is a result of excluding the interaction. The BIC is considerably more punitive than the AIC when it comes to adding parameters to the model. The reason that the additive model may be favoured over the interactive model, then, is not because the interactive model included the interaction term, but because it included more parameters than the additive model. In such circumstances, comparing a restricted model to the additive model may be useful. This restricted model may, for example, discard the main effect of concreteness, and retain the main effect of prime, and the two-way interaction between prime and concreteness. If it is then found that this restricted model is favoured over the additive model in the Bayes Factor and relative likelihood calculations, it can then be concluded that the data are more consistent with a model that contains the interaction term. In such circumstances, the fully interactive model should be chosen over the additive model, as it can be determined that the model that contains the interaction provides a better account of the data, and, unlike the restricted model, the results of the nonsignificant effects can still be reported.

Under circumstances where the Bayes factor and the relative likelihood favour different models, however, the relative values of the Bayes factor and relative likelihood were considered. For example, if the additive model is favoured over the restricted model with a Bayes factor of 2.00, but the restricted model is favoured over the additive model with a relative likelihood of 23.00, the restricted model would be selected, because the likelihood of the additive model resulting in a significant loss of information is far greater than the likelihood that the data are consistent with the assumptions of the additive model. In the circumstance where the additive model was favoured over the restricted model with a Bayes factor of 23.00, but the restricted model was favoured over the additive model with a relative likelihood of 2.00, the additive model would be selected, as the likelihood that the data are consistent with the assumptions of the additive model would be far greater than the likelihood that the use of the additive model would result in a significant loss of information.

2.2.3.1 Prime x sPIP Analysis

The model that was most favoured by the model selection analysis was one in which sPIP and the two-way interaction between sPIP and prime were included, indicating that while there was no overall main effect of prime on the results, the model selection favoured models in which the effect of prime interacted with subjects' sPIP score. This model was favoured over both the fully interactive model, $BF = 70.44$, $\theta = 2.17$, and the additive model, $BF = 9.97$, $\theta = 9.97$. Because there was almost 10 times more evidence for the restricted model than the additive model, this model selection analysis indicated that a model which includes the two-way interaction between prime and sPIP accounts for the data better than a model which excludes this interaction. The only reason that the additive model would be favoured over the fully interactive model, then, is due to the inclusion of prime as a fixed effect in the model. As such, the results are reported for the fully interactive model.

In the fully interactive model, there was no significant effect of prime, $t < 1$. Targets that were preceded by translation primes ($M = 651$ ms) produced the same RTs as targets that were preceded by control primes ($M = 652$ ms), replicating the results of prior research (e.g., Gollan et al., 1997). While the effect of sPIP was nonsignificant, $t < 1$, the two-way interaction between prime and sPIP was significant, $\beta = 3.56$, $SE = 1.58$, $t(7756) = 2.26$, $p = .024$. This two-way interaction is shown in Figure 1. The effect of prime on RTs, while not significant on its own, varied as a function of sPIP, with lower sPIP values being associated with an inhibitory effect of the prime, and larger sPIP values being associated with a facilitory effect of the prime relative to the control prime. Overall, subjects who reported higher listening and writing proficiency in English, as well as higher speaking and listening proficiency, but lower reading and writing proficiency in Chinese, produced larger priming effects.

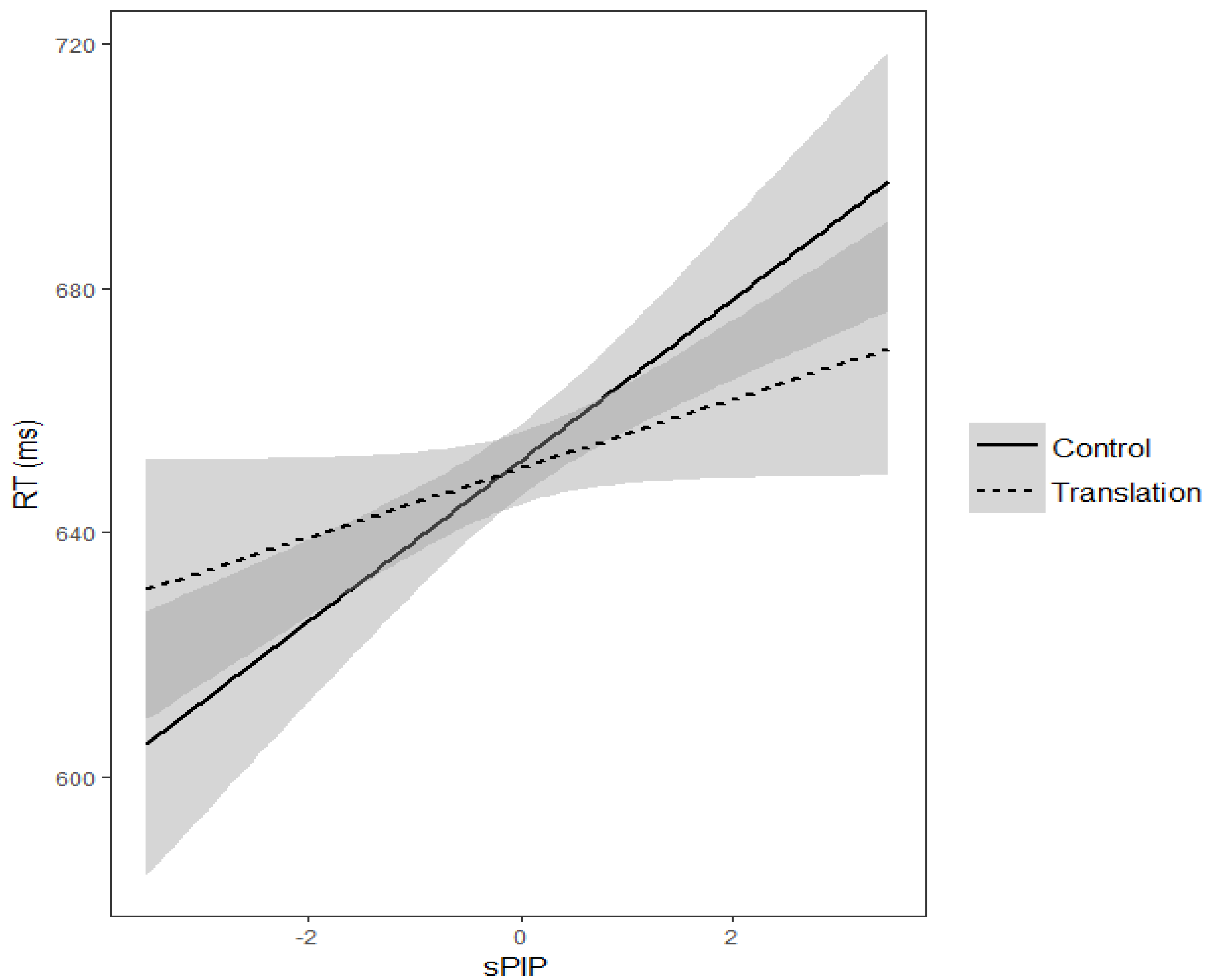

Figure 1. Response times as a function of prime and scaled sPIP, Experiment 1. Shaded areas represent 95% confidence intervals.

2.2.3.2 Prime x iPIP Analysis

As with the sPIP analysis, the restricted model was favoured over the full model, *BF* = 88.07, θ = 2.72, and the additive model, *BF* = 3.90, θ = 3.90. The main effect of prime was once again nonsignificant, $t < 1$, while the main effect of iPIP was significant, $\beta = -8.95$, $SE = 3.92$, $t(7756) = -2.29$, $p = .022$. The two-way interaction between iPIP and prime also approached significance, $\beta = 2.12$, $SE = 1.21$, $t(7756) = 1.76$, $p = .078$[3]. This two-way interaction is shown in Figure 2. As shown in Figure 2, larger iPIP scores were associated with faster RTs overall. The effect of prime was inhibitory for trials with a low iPIP score, and was facilitory for trials with a higher iPIP score. In sum, priming effects were larger when Chinese targets were lower in frequency, and when English primes were higher in frequency.

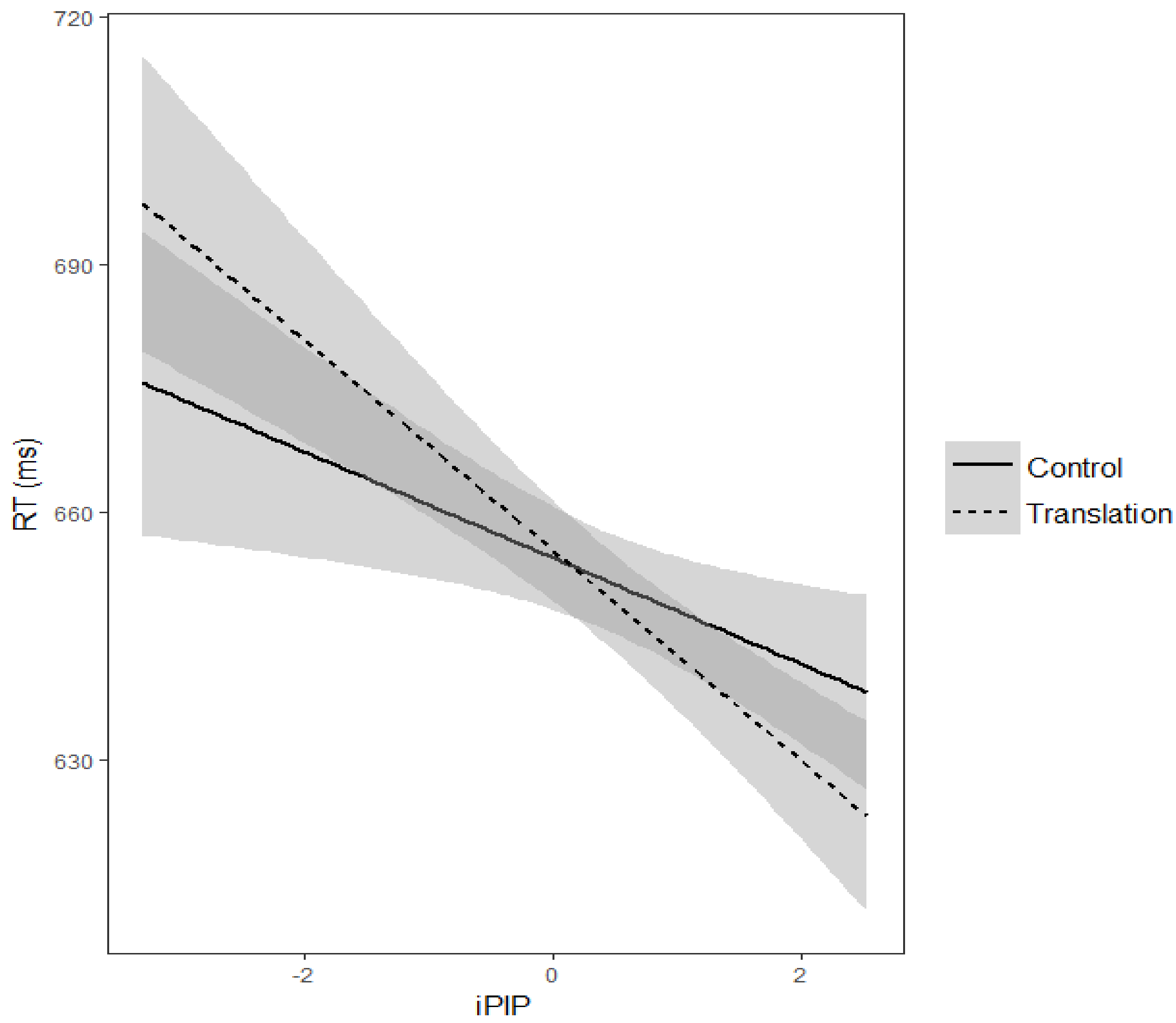

Figure 2. Response times as a function of prime and scaled iPIP, Experiment 1. Shaded areas represent 95% confidence intervals.

[3] The effect of iPIP was marginally significant when the criteria were loosened, $t(8678) = -1.80$, $p = .07$, and the two-way interaction was nonsignificant, $t(8596) = -1.28$, $p = .20$.

2.2.3.3 Prime x PIP Analysis

Once again, the restricted model was favoured over the fully interactive model, *BF* = 87.88, θ = 2.71, and the additive model, *BF* = 38.41, θ = 38.41. Neither the main effect of prime, $t < 1$, nor the main effect of PIP were significant, β = -9.30, *SE* = 6.12, $t(7756) = -1.52$, $p = .13$, but the two-way interaction between prime and PIP was significant, β = 3.82, *SE* = 1.41, $t(7756) = 2.72$, $p = .0065$. The two-way interaction between prime and PIP is shown in Figure 3. As shown in Figure 3, subject/item combinations with lower scores on PIP produced an inhibitory effect, while subject/item combinations with higher scores on PIP produced a priming effect. The effect of PIP changed as a function of prime type. For targets preceded by a translation prime, response times decreased as PIP increased. For targets preceded by a control prime, response times increased as PIP increased.

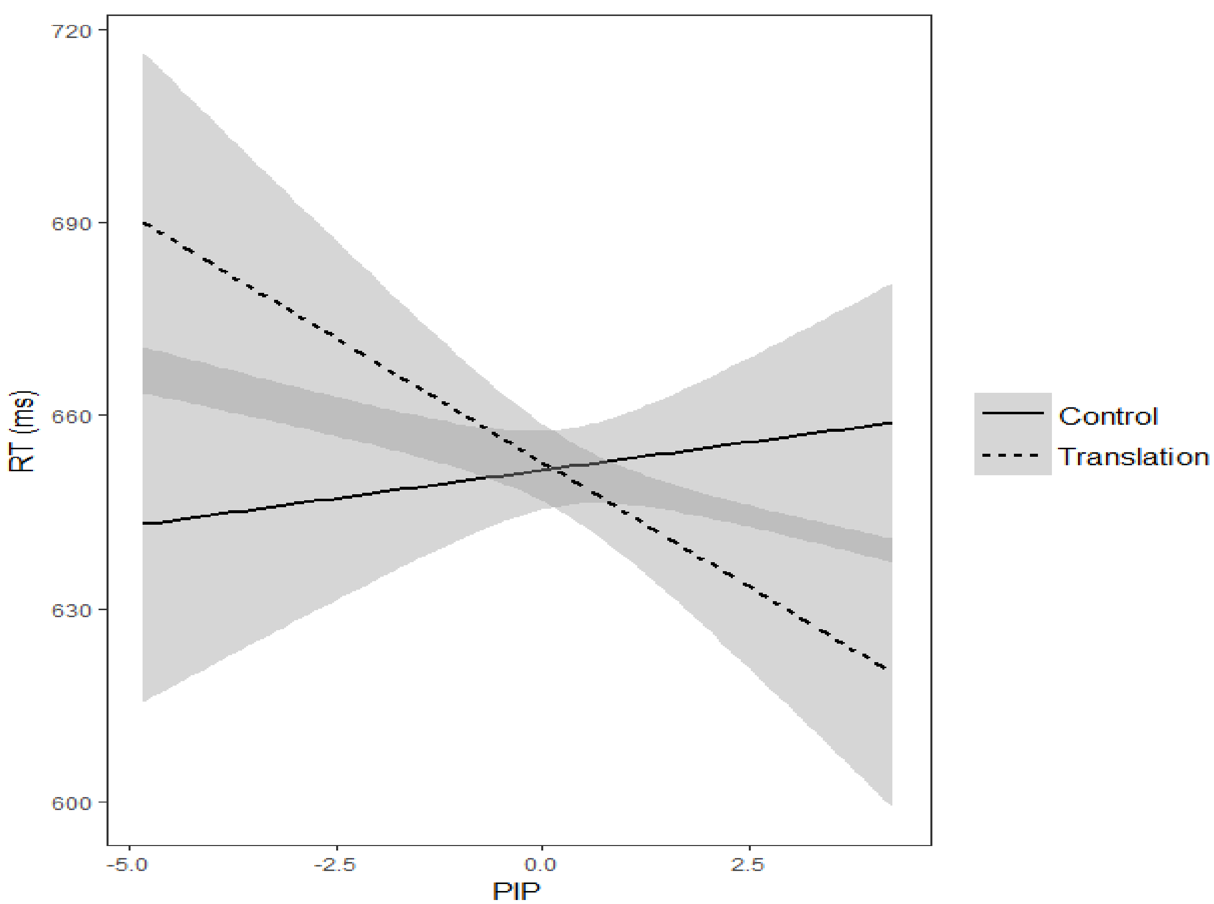

Figure 3. Response times as a function of prime and PIP, Experiment 1. Shaded areas represent 95% confidence intervals.

2.2.3.4 Prime x Experiment Order Analysis

To test whether the priming effect differed as a function of whether subjects finished Experiment 1 or Experiment 2 first, a follow-up analysis was conducted using prime and experiment order as fixed effects. This analysis found no significant effects of prime, experiment order, nor a two-way interaction between prime and experiment order, ts < 1. Numerically, latencies were shorter for participants who completed Experiment 2 (M = 648 ms) before Experiment 1 (M = 655 ms), but this difference was nonsignificant. When subjects did Experiment 1 before Experiment 2, trials that were preceded by a control prime (M = 656 ms) produced similar latencies to trials that were preceded by translation primes (M = 654 ms). Likewise, when subjects did Experiment 2 before Experiment 1, trials that were preceded by a control prime (M = 649 ms) produced similar latencies to trials that were preceded by translation primes (M = 647 ms).

2.2.3.5 Prime x List Analysis

To test whether the priming effect was affected by the counterbalance list used, a follow-up analysis was conducted using prime and counterbalance list as fixed effects. This analysis found no significant effects of prime, list, nor a two-way interaction between prime and list, ts < 1. Numerically, in List 1, there was an inhibitory effect of prime (-5 ms), while in List 2, there was a small advantage for primes (8 ms)[4]. The interaction was not significant, however[5].

2.2.4 Error Analysis

2.2.4.1 Prime x sPIP Analysis

The error data were separately submitted to a generalized linear mixed effects model using a binomial distribution. Error models were fit without the use of random slopes for any analyses. The model most favoured by the model selection analysis was the additive model, which was

[4] Neither the inhibitory effect, nor the facilitative effect were significant when each list was analyzed in isolation, ts < 1.

[5] When assessing possible reasons why the priming effects were slightly different in Lists 1 and 2, the mean sPIP and iPIP characteristics were compared for positive trials that were preceded by a translation prime. None of the sPIP factors differed significantly between lists, ts < 1.50, ps > .13. For the iPIP factors, however, there was a difference between the prime frequency of items in List 1 (M = 72) and List 2 (M = 82), $t(3871) = -3.86$, p = .0001.

favoured over the fully interactive model, *BF* = 88.21, θ = 3470598072. The effect of prime was not significant, β = -0.11, *SE* = 0.08, *z*(8064) = -1.30, *p* = .19. Targets that were preceded by a control prime (*M* = 2.26%) produced identical error rates to targets that were preceded by a translation prime (*M* = 2.13%). Neither the effect of sPIP, *t* < 1, nor the two-way interaction were significant, *t* < 1. Mean error rates as a function of sPIP tertile can be found in Figure 4.

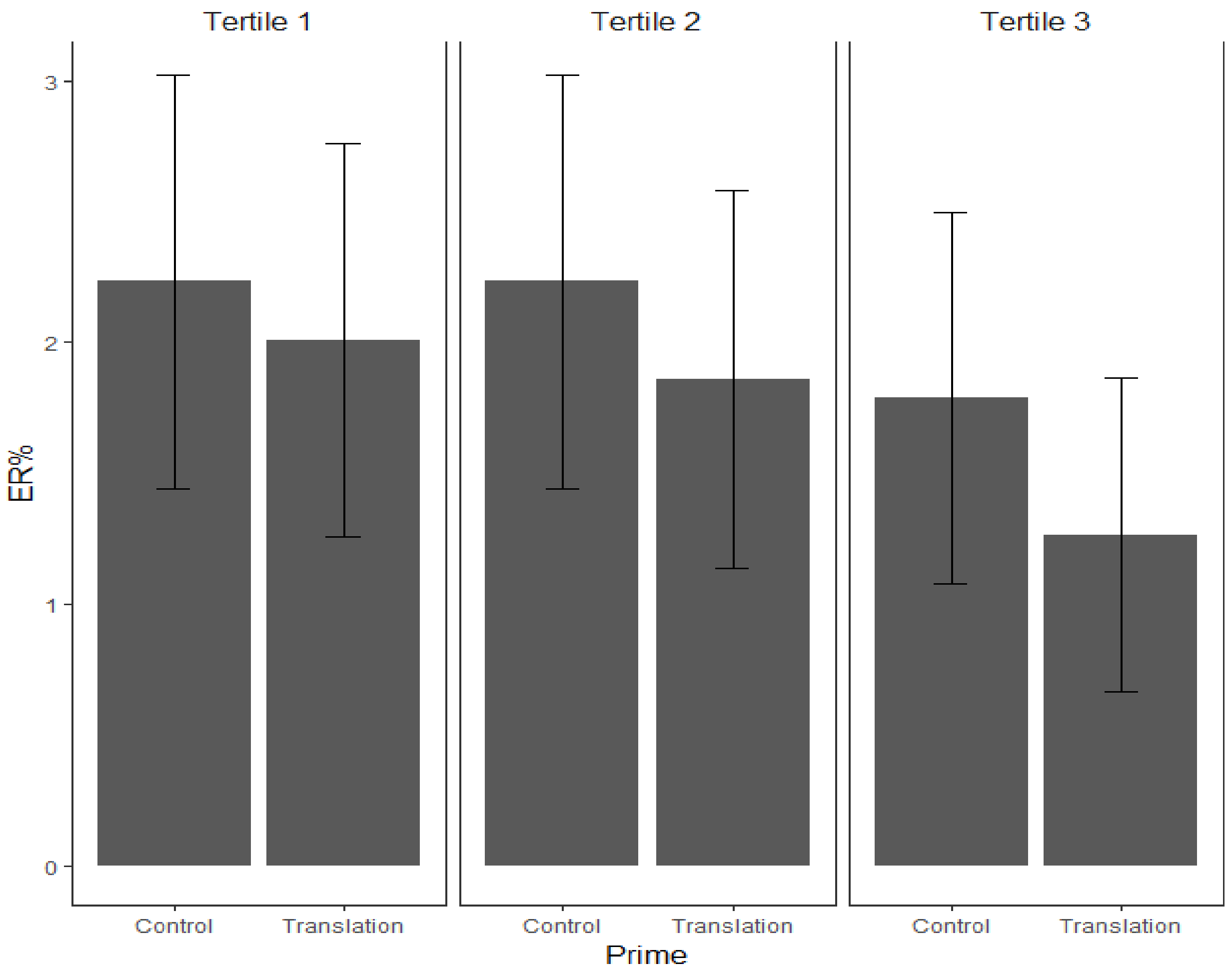

Figure 4. Error rates as a function of prime and sPIP tertile, Experiment 1. Error bars represent 95% confidence intervals.

2.2.4.2 Prime x iPIP Analysis

The restricted model which included the main effect of iPIP and the two-way interaction between prime and iPIP was favoured over both the fully interactive model, *BF* = 32.53, θ = 0.98, and the additive model, *BF* = 4.54, θ = 4.54. The main effect of prime was nonsignificant, β = -0.12, *SE*

= 0.09, $z(8064) = -1.42$, $p = .16$, while the main effect of iPIP, $\beta = 0.23$, $SE = 0.09$, $z(8064) = 2.45$, $p = .014$, and the two-way interaction between prime and iPIP were significant, $\beta = 0.13$, $SE = 0.06$, $z(8064) = 2.22$, $p = .027$[6]. This two-way interaction is shown in Figure 5. As seen in Figure 5, items in Tertiles 1 and 2 of the iPIP score produced identical error rates for items preceded by control and translation primes, while items in Tertile 3 produced a priming effect in the error rates, which was largely driven by slightly higher error rates in the control condition (2.8%) than in the translation condition (1.7%).

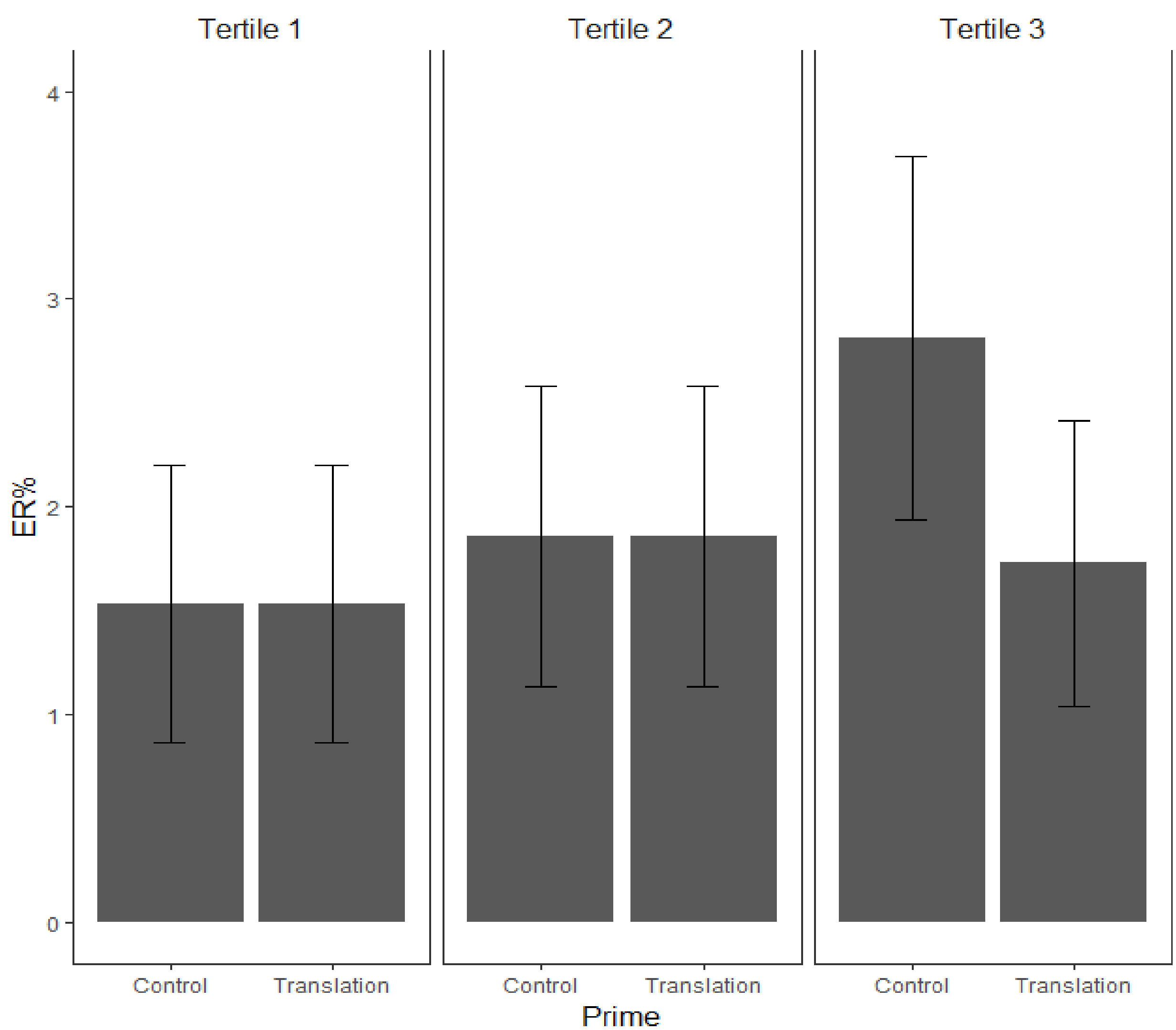

Figure 5. Error rates as a function of prime and iPIP tertile, Experiment 1. Error bars represent 95% confidence intervals.

[6] This two-way interaction was nonsignificant when the screening criteria were loosened, $z < 1$.

2.2.4.3 Prime x PIP Analysis

The restricted model which included the main effect of PIP and the two-way interaction between prime and PIP was favoured over both the fully interactive model, *BF* = 28.84, $\theta = 0.87$, and the additive model, *BF* = 2.05, $\theta = 2.05$. The effects of prime, $\beta = -0.13$, SE = 0.085, $z(8064) = -1.50$, $p = .13$, and PIP were nonsignificant, $\beta = 0.16$, $SE = 0.10$, $z(8064) = 1.52$, $p = .13$, while the two-way interaction between prime and PIP approached significance, $\beta = 0.12$, $SE = .07$, $z(8064) = 1.77$, $p = .076$[7]. The mean error rates as a function of PIP tertile and prime are shown in Figure 6. As seen in Figure 6, Tertiles 1 and 2 produced a null effect of the prime on error rates, while the difference between targets preceded by control primes (1.93%) and translation primes (1.27%) in Tertile 3 was marginally significant.

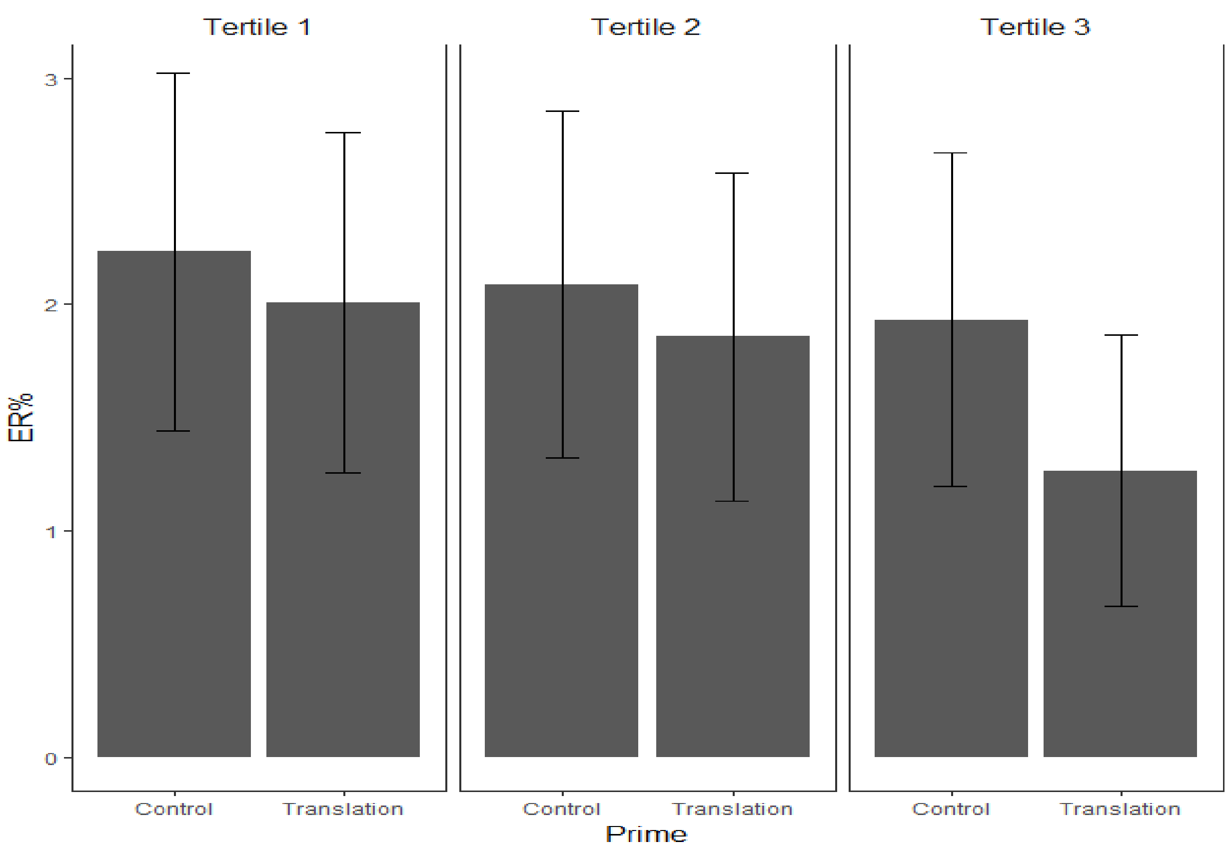

Figure 6. Error rates as a function of prime and PIP tertile, Experiment 1. Error bars represent 95% confidence intervals.

[7] The two-way interaction was nonsignificant when the screening criteria were loosened, $z < 1$.

2.3 Discussion

It was initially hypothesized that priming in the LDT could be predicted as a function of subjects' skill levels across different domains of L2 proficiency. Specifically, it was predicted that domains of English proficiency associated with orthographic coding should be associated with priming effects in the LDT. Using measures from models of machine learning to derive a set of feature weights for how an array of factors impact priming effects in the LDT, Experiment 1 has provided tentative support for this prediction, but with the caveat that factors such as the verbal comprehension of English are highly important. Subjects' self-reported writing ability in English was one of the strongest predictors of priming among measures examined, with the results showing evidence that priming is impacted by expressive writing abilities in L2. Critically, however, the effects of prime and sPIP – as created using positive factors such as Chinese speaking and listening, and English listening and writing abilities, and negative factors such as Chinese reading and writing abilities – were null when examined in isolation. Primes had little impact on RTs in the overall data, which replicated the results of prior studies (e.g., Finkbeiner et al., 2004; Gollan et al., 1997; Grainger & Frenck-Mestre, 1998; Jiang & Forster, 2001), as did the sPIP score. It was only through the combination of these factors that they significantly affected RTs in lexical decision. Priming effects were facilitative for subjects who reported higher speaking and listening proficiency in Chinese, and listening and writing abilities in English, but weaker reading and writing abilities in Chinese. For subjects who were less proficient at writing and comprehending spoken English, and who were stronger readers and writers in Chinese, priming effects were inhibitory.

The fact that facilitative priming effects were larger for L2 learners who reported weaker reading and writing abilities in their L1 shouldn't come as a surprise. Similar results have been reported in studies in the L1-L2 direction by Nakayama et al. (2012, 2013), who found that L1-L2 priming effects were larger when subjects were less proficient in their L2 than when they were more proficient in their L2. Further, for subjects who are less skilled or experienced with their L1 orthographic system, tasks that emphasize lexical orthographic knowledge, such as the lexical decision task, would be more burdensome for them. In such cases, the processing of targets is less efficient, reducing the likelihood that a floor effect would occur, and would provide more opportunity for the prime to influence the decision. What is critical is that subjects are also familiarized and skilled with their L2 orthographic system, as the knowledge and familiarity of word forms in a learner's L2 would be a good indicator that the subject not only knows the L2 words that are used as primes, but also that their knowledge of the word's meaning has also been sufficiently bound to the form representations of the prime. Without this necessary knowledge of L2 word form and meaning, the prime will not sufficiently activate the target, and it is more likely to facilitate decisions once the target meaning is activated if the subject is not as skilled with their L1 orthography.

In addition, Experiment 1 also tested whether item-specific factors such as prime and target frequency would impact the priming effects obtained in lexical decision. The evidence that item-based factors affected L2-L1 priming were mixed. Experiment 1's results showed that facilitative effects were more likely to occur when the primes are higher in frequency, and the targets are lower in frequency. Such a finding is consistent with an account that assumes that the prime's ability to preactivate the target's meaning is dependent on the resting-level activation of the prime's word representation. Primes with higher resting-level activations are more likely to preactivate the target than primes with lower resting-level activity. Likewise, there is more opportunity for the prime to facilitate processing on the target when the resting-level activation of the target is lower. One circumstance where the resting-level activity of the target would be lower is when the target is low-frequency. The latter result showing that priming effects were larger for low-frequency targets is again consistent with Nakayama et al.'s (2012, 2013) results showing that priming effects were larger for low-frequency targets than high-frequency targets in L2-L1 translation priming, and expands on these studies, showing that priming effects are larger when high-frequency primes are used than when low-frequency primes are used. However, these findings were only found with more stringent outlier screening. When the criteria for outlier screening were loosened, the two-way interaction between prime and iPIP was no longer significant. While the loss of this interaction may be due to a larger number of outlier data included in the analysis, these results suggest that if these factors have an influence on processing, it is a weak effect.

One final issue to be noted is that the results of Experiment 1 suggest that under certain circumstances, the prime has an inhibitory effect on processing[8]. Any explanation as to why translation primes would produce a priming effect is speculation, as there doesn't appear to be any obvious explanation for why this inhibition occurred. What this inhibition may suggest is that subjects who are less proficient in their L2 still process the prime to an extent, but the processing of such primes and the attempted retrieval of meaning-level information associated with the target is highly inefficient at lower proficiencies, and comes at a cost when compared to responding to targets that were preceded by an unrelated prime. In such circumstances, no additional processing of the control prime is engaged, requiring less resources to be allocated to it. As a result, the translation prime produces inhibition, rather than facilitation.

A more comprehensive account of these findings will be presented in the General Discussion. Before doing so, I will turn to a second issue, whether there is evidence of a dissociation in the L2 skills and item-specific factors that predict priming in lexical decision and semantic categorization. As was noted previously, priming effects in the SCT, unlike in the LDT, may be affected by the amount of time that subjects use their L2 in their daily lives, across different social environments. Specifically, it is possible that subjects who use their L2 at home, at school, and in other social contexts more frequently should have more opportunities to acquire a richer base of semantic knowledge associated with their L2. The semantic categorization task, while still requiring sufficient L2 orthographic knowledge, should not place as much emphasis on this knowledge as it does on the development of L2 semantic knowledge, as obtained through the use of the language in naturalistic social settings. Further, because semantic categorization is less sensitive to frequency-based information than the lexical decision task, it might be expected that the importance of prime frequency in this task would be diminished compared to lexical decision. In semantic categorization, priming might be predicted to be less dependent on the specification of L2 word representations, so long as the prime activates information about the target's category membership. These ideas were examined in Experiment 2.

---

[8] This effect was consistent in the sPIP data, but not the iPIP data, when changing the screening criteria.

Chapter 3

3 Experiment 2

3.1 Method

3.1.1 Subjects

Subjects were the 103 subjects who had also participated in Experiment 1.

3.1.2 Stimuli

Experiment 2 consisted of 200 trials across five blocks of 40 trials, with 20 exemplars and 20 nonexemplars of a selected category in each block. Five categories were used for the exemplars and nonexemplars: mammals, insects, body parts, vegetables/fruits, and clothing/accessories. Each word appeared twice in the experiment, appearing as an exemplar in one block, and as a nonexemplar in another block. Nonexemplars in each block were taken from four of the other categories, with five nonexemplars taken from each category. Half of the exemplars and nonexemplars were preceded by a translation prime, while the other half was preceded by a control prime. For both exemplars and nonexemplars, control primes were from a different semantic category than the target. Primes were counterbalanced across two lists, such that each target appeared with a translation and a control prime once across both lists. Mean ratings for stroke count and frequency for all targets can be found in Table 2. None of the stimuli that appeared in Experiment 2 appeared in Experiment 1. A list of all of the stimuli used in Experiment 2 is found in Appendix C.

3.1.3 Measures

The measures were identical to the measures that were used in Experiment 1, with the exception that category typicality ratings were included based on the prime language data. Category typicality ratings were derived from three separate sources: Rosch's (1975) norms, and Uyeda and Mandler's (1980) norms. Because neither of these norms provided data on the mammal or insect categories, additional data on category typicality had to be derived from Ruts et al.'s

(2004) norms in Dutch. Despite the Ruts et al. norms being in Dutch, it was deemed that there was enough cultural overlap that these norms would provide a reasonably accurate assessment of the typicality of insects and mammals in English[9]. Due to the scale of these ratings differing, the typicality values in each of the norms were rescaled to ensure that all data were using the same scope of values.[10] Otherwise, the only difference between Experiment 2 and Experiment 1 was that the PIP was now fit using the priming data from the semantic categorization task.

3.1.4 Procedure

Experiment 2 was completed in the same session as Experiment 1. Subjects first entered the lab and were greeted by the experimenter. After reading through a letter of information and obtaining informed consent, subjects then completed the LEQ as thoroughly as possible. Subjects were then seated in front of a computer. Subjects were instructed to indicate whether each target was a member of a target category or not as quickly and as accurately as possible by pressing either the right shift key for exemplars, or the left shift key for nonexemplars. Subjects initially received 8 practice trials before beginning the experiment, in which the target category was weapons. After the practice trials, a new set of instructions was presented, allowing the subjects to take a break, and informing them what the target category was going to be for the next block. The order of block presentation was counterbalanced, and the order of trials within each block was randomized. The set of instructions for each block was always set up in a way that it was paired with the correct block. For example, the instructions denoting that the target category is mammals would always appear with the block in which the exemplars were mammals, the instructions denoting that the target category is fruits/vegetables always appeared with the block containing fruit/vegetable exemplars, etc. Subjects completed five of these blocks of 40 stimuli and were always given a break with a new set of instructions about the new target category after

[9] There was a significant correlation between the typicality ratings for items that appeared in Uyeda and Mandler's (1980) English norms and the Ruts et al. (2004) Dutch norms, $r(38) = -.69$, $p < .0001$. The correlation was negative because smaller scores in Uyeda and Mandler's norms denoted more typical category members, while smaller scores in the Ruts et al. norms denoted more atypical category members.

[10] Typicality ratings were included post-hoc, after data were collected. Typicality ratings were not available for all stimuli. Data are first reported without the typicality ratings. The effects of typicality are reported in the combined data from Experiments 2 and 3.

the block. Upon completing both the SCT and LDT, the subjects were then debriefed, and were then dismissed.

3.2 Results

3.2.1 Data Trimming

The data for Experiment 1 and Experiment 2 were trimmed simultaneously. Information on the trimming procedure can be found in the Results section for Experiment 1.

3.2.2 PIP

The coefficients for the model derived for Experiment 2 can be found in Table 7. With respect to the subject-based predictors, the largest predictors of priming effects in the SCT were the percentage of time that subjects used English in the school environment, their self-reported speaking proficiency, and the percentage of time that subjects used English in social environments outside of home and school. Negatively associated with priming effects were self-reported writing and speaking proficiency in Chinese. With respect to the item-based predictors, the largest predictors of priming effects in the SCT were target frequency and the number of strokes. Prime frequency, while still positively associated with priming effect sizes, had a reduced impact.

Table 7. *PIP Coefficients for Experiments 2 & 3.*

	PIP Coefficient Values
sPIP	
CW	-2.18
CS	-1.59
CR	-1.26
EL	0.81
CL	1.62
PEO	3.11
ES	5.03
PES	9.21
iPIP	
GF	6.87
L	1.53
PCEL	0.78

Note: CW = Self-rated Chinese writing proficiency; CS = Self-rated Chinese speaking proficiency; CR = Self-rated Chinese reading proficiency; EL = Self-rated English listening proficiency; CL = Self-rated Chinese listening proficiency; PEO = Percentage of time English is used in other social settings; ES = Self-rated English speaking proficiency; PES = Percentage of time English is used at school; GF = Target Google frequency; L = Prime length; PCEL = Prime CELEX frequency.

3.2.3 Reaction Time Analysis.

3.2.3.1 Prime x sPIP Analysis

The raw response times for exemplar trials were submitted to a generalized linear mixed effects model using the lme4 package (Bates et al., 2015) using R (R Core Team, 2017). The model included prime type and sPIP as fixed effects, and subjects and items as random effects. The model was fit using an inverse Gaussian distribution. The relationship between prime and sPIP is shown in Figure 7. For all analyses in Experiment 2, the results for nonexemplars are described and shown in Appendix F.

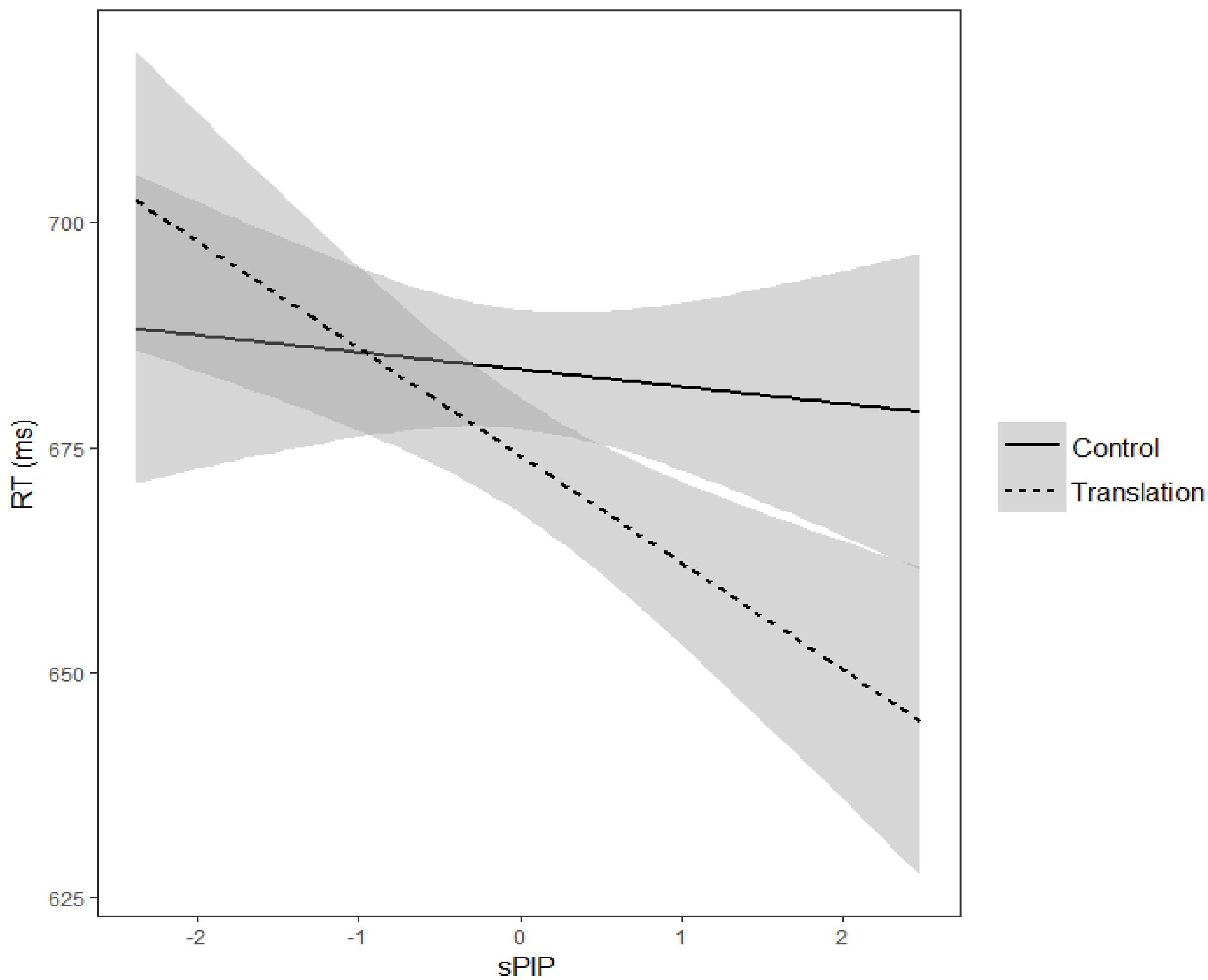

Figure 7. Response times as a function of prime and scaled sPIP, Experiment 2 exemplars. Shaded areas represent 95% confidence intervals.

The model selection analysis produced mixed results. The model that was most favoured by the Bayes Factor analysis was the additive model, which outperformed both the full model, *BF* = 10.59, and a restricted model that excluded the main effect of sPIP, but retained all of the interactions, *BF* = 10.60. However, the fully interactive model outperformed the additive model in the relative likelihood analysis, θ = 2.97, and performed similarly to the restricted model, θ = 1.00.

The additive model involved a significant effect of prime, $\beta = -10.04$, $SE = 3.50$, $t(7297) = -2.87$, $p = .0042$. Targets preceded by a translation prime ($M = 674$ ms) produced faster latencies than targets preceded by a control prime ($M = 684$ ms), replicating the translation priming effect found in prior research (e.g., Grainger & Frenck-Mestre, 1998). There was no effect of sPIP, $\beta = -10.15$, $SE = 9.30$, $t(7297) = -1.09$, $p = .27$. The interactive model additionally involved a marginally significant two-way interaction between sPIP and prime type, $\beta = -13.83$, $SE = 8.32$, $t(7297) = -1.66$, $p = .097$[11]. In general, subjects who reported using English a larger proportion of time at school and in social settings outside of home and school, reported higher speaking and listening proficiency in English, higher listening proficiency, but lower writing, speaking, and reading proficiency in Chinese produced larger priming effects.

3.2.3.2 Prime x iPIP Analysis

Once again, the model selection results were mixed. For the Bayes Factor, the additive model was favoured over the full model, $BF = 12.02$, and the restricted model which retained the effects of prime and the two-way interaction between prime and iPIP, $BF = 12.02$. However, for the relative likelihood, the full model was favoured over the additive model, $\theta = 2.61$. For the additive model, the main effect of prime was significant, $\beta = -9.70$, $SE = 3.48$, $t(7297) = -2.79$, $p = .005$, while the effect of iPIP was not, $\beta = -2.97$, $SE = 2.96$, $t(7297) = -1.00$, $p = .32$. In the interactive model, the two-way interaction between prime and iPIP was significant, $\beta = -8.50$, $SE = 4.14$, $t(7297) = -2.05$, $p = .04$[12]. This interaction is shown in Figure 8. As can be seen in Figure increased for targets preceded by a translation prime, and RTs increasing as iPIP increased for targets preceded by a control prime. Priming effects were larger when the targets were higher frequency, and were also impacted by the prime length, with targets preceded by longer primes 8, the effect of iPIP on RTs varied as a function of prime type, with RTs decreasing as iPIP producing larger priming effects than targets preceded by shorter primes.

[11] The two-way interaction was statistically significant when the screening criteria were loosened, $t(8354) = -2.55$, $p = .011$.
[12] The two-way interaction increased when the screening criteria were loosened, $t(8354) = -2.80$, $p = .005$.

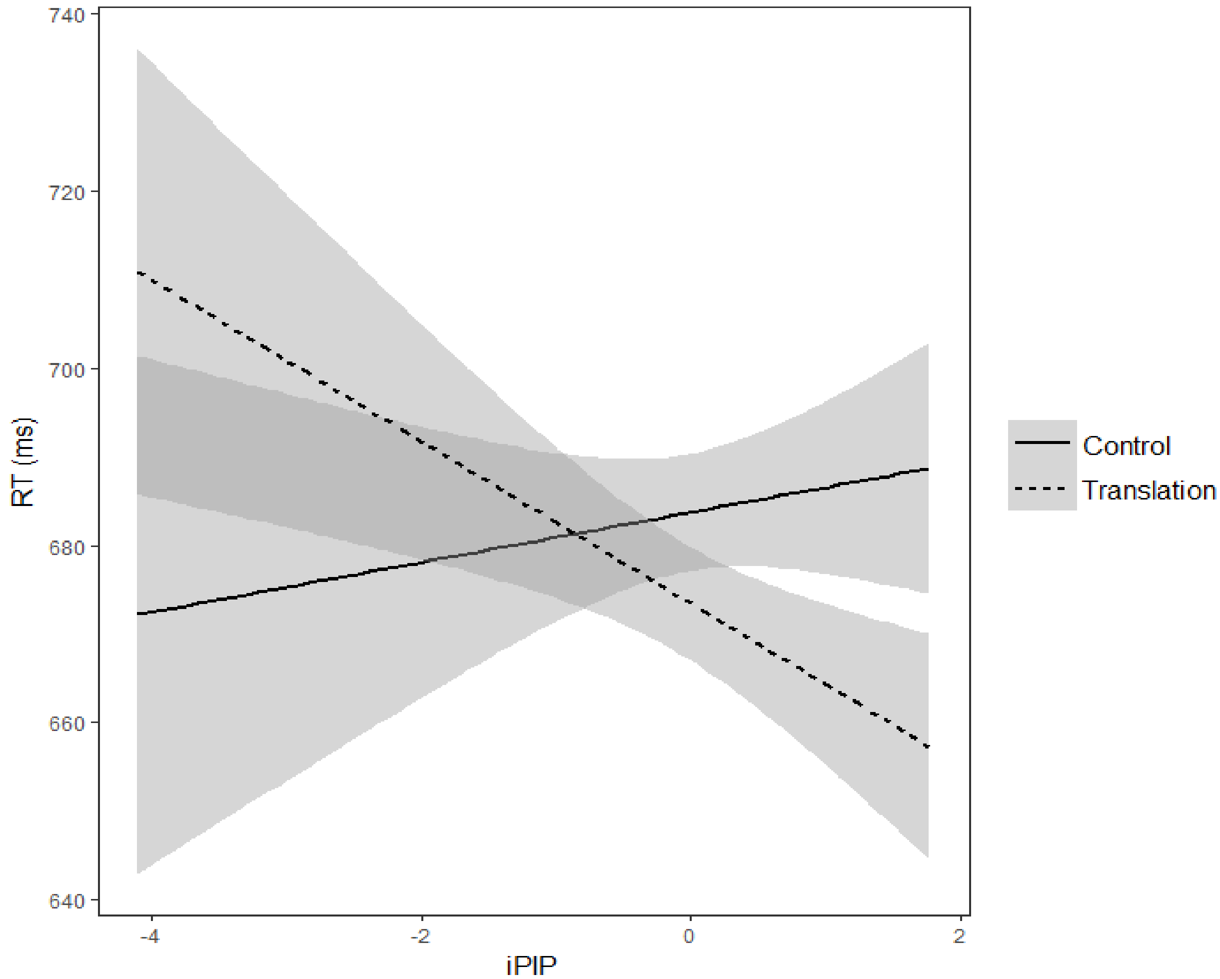

Figure 8. Response times as a function of prime and scaled iPIP, Experiment 2 exemplars. Shaded areas represent 95% confidence intervals.

3.2.3.3 Prime x PIP Analysis

As with the sPIP and iPIP analyses, the Bayes Factor favoured the additive model over both the full model and the restricted model, *BF* = 1.45, but the relative likelihood favoured both the full model, θ = 21.63, and the restricted model, θ = 21.66, over the additive model. In these analyses, the main effect of prime was significant, β = -9.45, *SE* = 3.59, *t*(7297) = -2.63, *p* = .0085, while the main effect of PIP was not, *t* < 1. The two-way interaction between prime and PIP was significant, β = -11.53, *SE* = 3.78, *t*(7297) = -3.05, *p* = .0023[13]. This interaction is seen in Figure 9. As can be seen in Figure 9, the effect of PIP on RTs once again varied as a function of prime

[13] The two-way interaction increased when the screening criteria were loosened, *t*(8354) = -3.62, *p* = .0003.

type. For targets preceded by translation primes, increases in PIP were associated with faster response times. For targets preceded by control primes, PIP had no effect on response times. The result was that the priming effect grew larger as the PIP score increased, demonstrating that the combination of subject- and item-specific factors used to compute the sPIP and iPIP scores significantly predicted priming effects in Experiment 2.

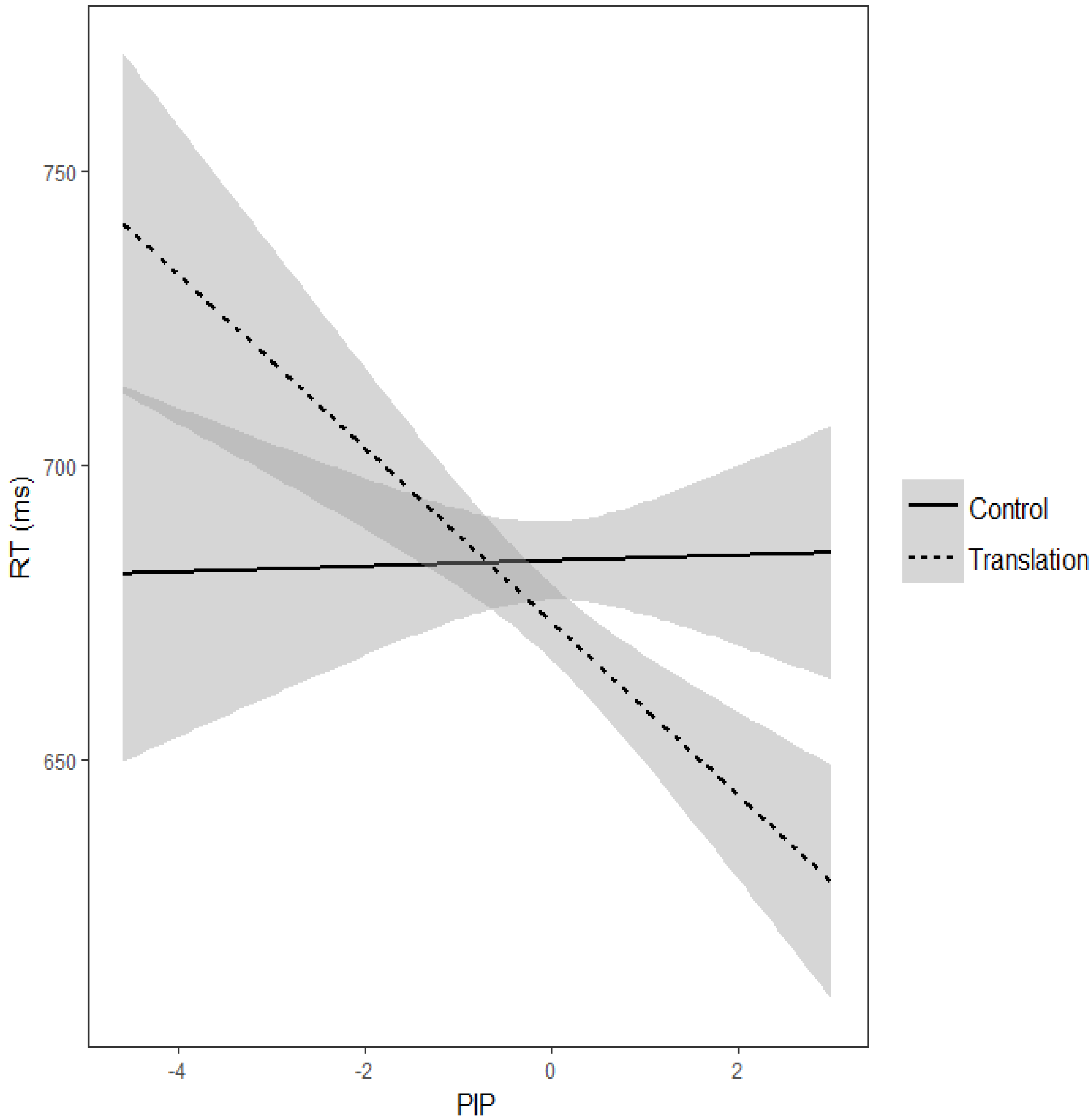

Figure 9. Response times as a function of prime and PIP, Experiment 2 exemplars. Shaded areas represent 95% confidence intervals.

3.2.3.4 Prime x Experiment Order Analysis

Regardless of what model was selected, only the effect of prime was significant, $\beta = -14.12$, $SE = 4.36$, $t(7297) = -3.24$, $p = .001$. Neither the effect of experiment order nor the two-way interaction was significant, $t < 1$[14]. Response times were virtually identical when subjects completed Experiment 1 before Experiment 2 ($M = 683$ ms) compared to when subjects completed Experiment 2 before Experiment 1 ($M = 675$ ms). Numerically, the priming effect was larger when subjects completed Experiment 1 before Experiment 2 (14 ms) than when they completed Experiment 2 before Experiment 1 (6 ms), but this difference was nonsignificant.

3.2.3.5 Prime x List Analysis

The only effect that was found to be significant was the effect of prime, $\beta = -24.58$, $SE = 8.19$, $t(7297) = -2.99$, $p = .003$. Neither the effect of list, nor the two-way interaction were significant, *t*s < 1. The priming effect in List 1 (11 ms) was not significantly different from the priming effect in List 2 (8 ms).

3.2.4 Error Analysis

3.2.4.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, $BF = 83.69$, $\theta = 3123994119$. The main effect of prime was nonsignificant, $z < 1$[15]. Targets preceded by translation primes ($M = 4.82\%$) produced similar error rates to targets that were preceded by control primes ($M = 5.22\%$). There was a significant effect of sPIP on error rates, $\beta = 0.20$, $SE = 0.08$, $z(7896) = 2.42$, $p = .016$[16], but a nonsignificant two-way interaction in the fully interactive model, $z < 1$. The effects of prime and sPIP tertile on error rates is shown in Figure 10. As shown in Figure 10,

---

[14] The two-way interaction was marginally significant when the screening criteria were loosened, $t(8354) = -1.76$, $p = .079$. The priming effect was larger when participants completed Experiment 1 first (21 ms) than when they completed Experiment 2 first (10 ms).

[15] The effect of prime was nonsignificant when the criteria were loosened, $z(9025) = 1.28$, $p = .20$.

[16] The effect of sPIP was nonsignificant when the criteria were loosened, $z(9025) = 1.28$, $p = .20$.

error rates in Tertile 3 (*M* = 6.69%) were larger than they were in Tertiles 1 (*M* = 4.67%) and 2 (*M* = 3.69%).

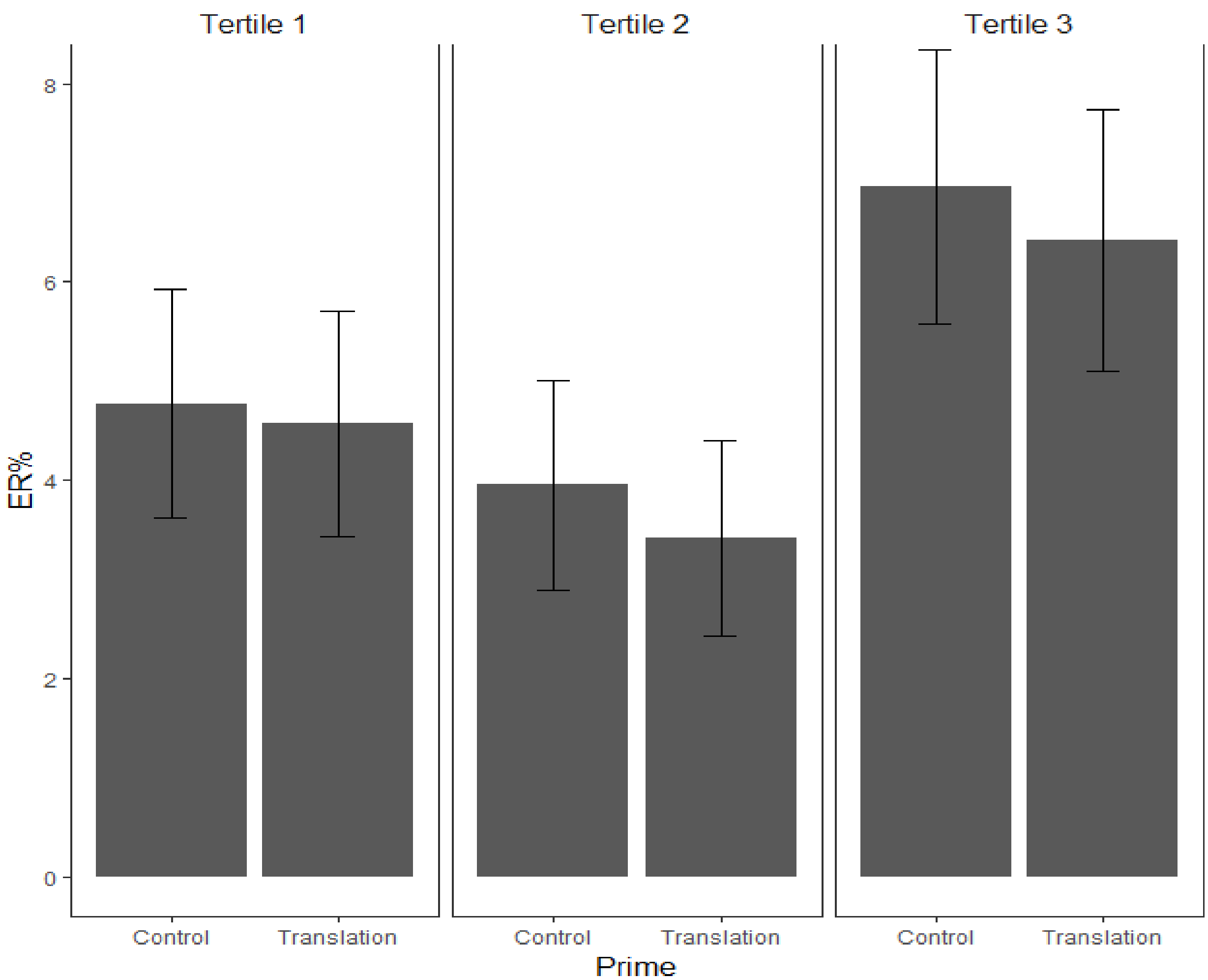

Figure 10. Error rates as a function of prime and sPIP tertile, Experiment 2 exemplars. Error bars represent 95% confidence intervals.

3.2.4.2 Prime x iPIP Analysis

The additive model was again favoured over the full model, *BF* = 63.88, θ = 2384754990, but none of the effects or interactions were significant, all *z*s < 1. The effects of prime and iPIP tertile on mean error rates are shown in Figure 11. Trials in Tertile 1 (*M* = 3.17%), Tertile 2 (*M* = 3.73%), and Tertile 3 (*M* = 3.27%) produced similar error rates, and there was no difference in the effect of the prime on error rates across tertiles.

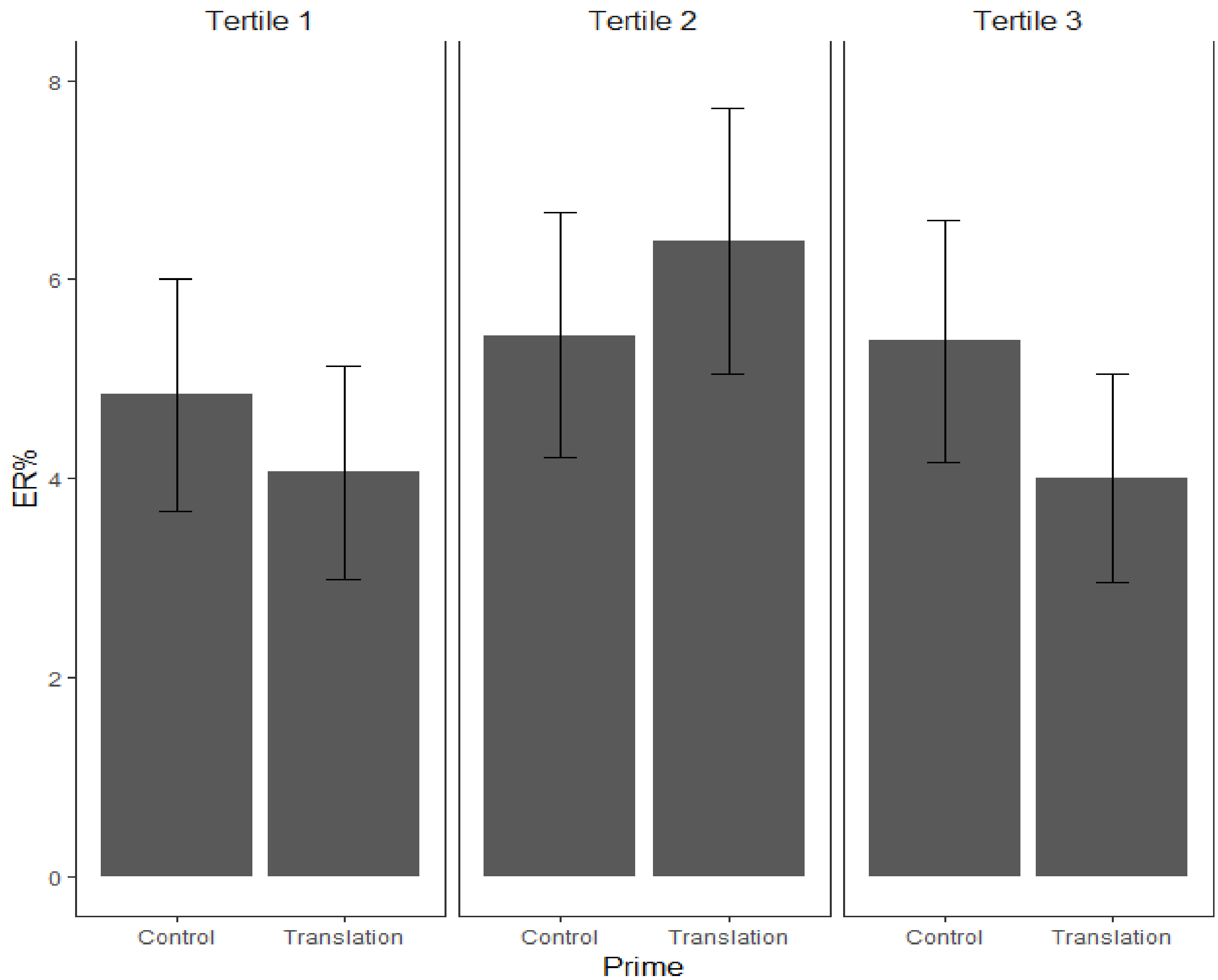

Figure 11. Error rates as a function of prime and iPIP tertile, Experiment 2 exemplars. Error bars represent 95% confidence intervals.

3.2.4.3 Prime x PIP Analysis

The additive model was favoured over the full model, *BF* = 62.24, θ = 2323376759. The only effect that trended in this analysis was the effect of PIP, which approached significance, $\beta = 0.15$, $SE = 0.08$, $z = 1.94$, $p = .052$[17]. All other effects and interaction terms were nonsignificant, *z*s < 1. The mean error rates as a function of prime and PIP tertile for Experiment 2 exemplars are shown in Figure 12. Overall, there was a difference between error rates in Tertile 3 ($M = 6.88\%$)

[17] The effect of PIP was nonsignificant when the criteria were loosened, $z(9025) = 1.51$, $p = .13$.

and Tertiles 1 (M = 4.60%) and 2 (M = 3.57%). More importantly, there was no difference in the priming effects among the Tertiles.

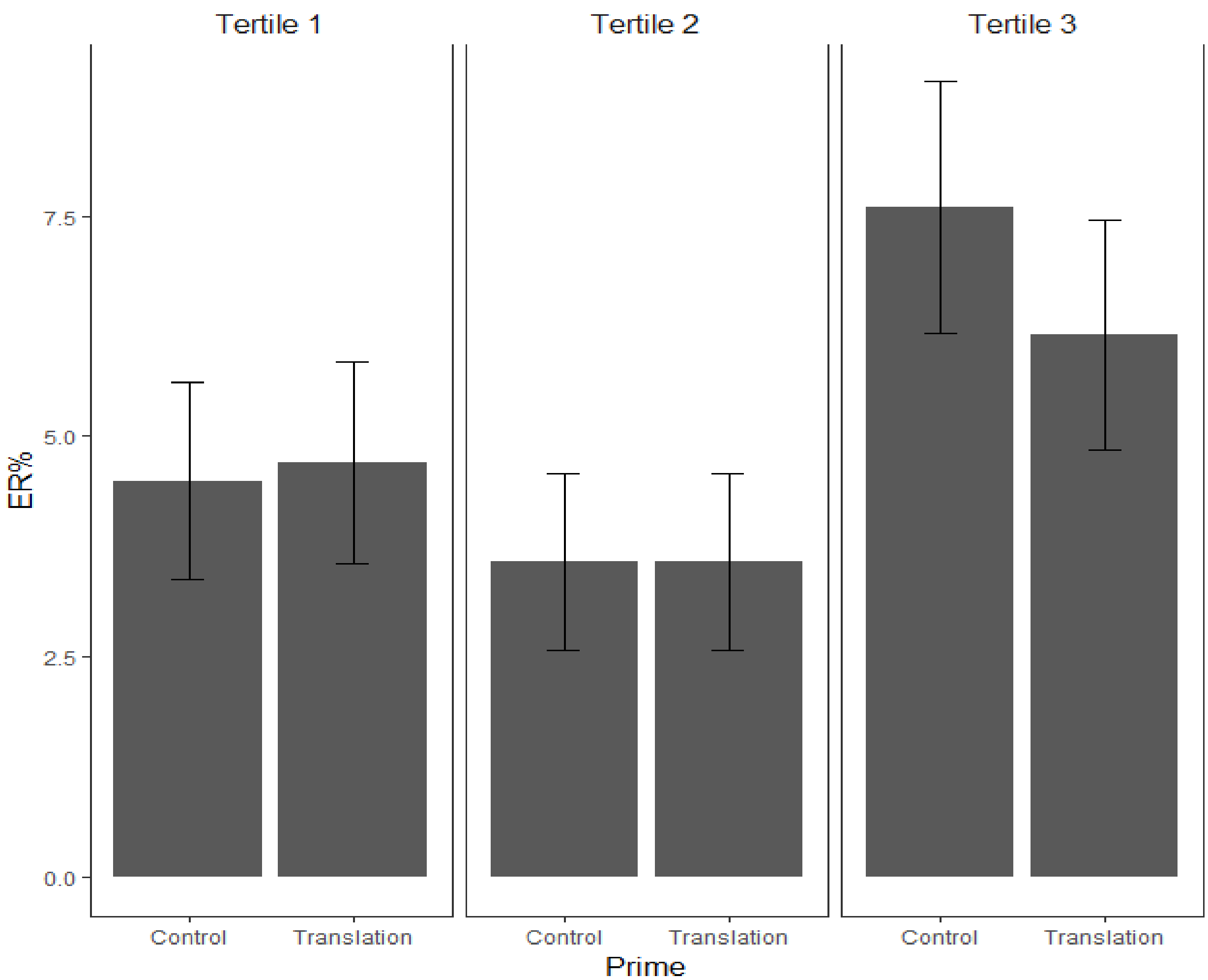

Figure 12. Error rates as a function of prime and PIP tertile, Experiment 2 exemplars. Error bars represent 95% confidence intervals.

3.3 Discussion

An initial expectation going into Experiment 2 was that there would be evidence of association between the L2 skills, behaviours, and item-specific factors that would predict priming effects in an LDT and the skills, behaviours, and item-specific factors that would predict priming effects in an SCT. Specifically, priming in the LDT in Experiment 1 was associated with productive writing abilities in one’s L2, and was highly sensitive to the frequency of the prime. Priming in an SCT, on the other hand, was predicted to be affected by the amount of time subjects use their L2 across a broad array of social contexts, as using their L2 would provide more opportunities to acquire a greater breadth and depth of semantic knowledge of words in their L2, and indicates that subjects are more immersed in the English-speaking social environment. The importance of prime frequency, in such cases, should be reduced, as the only information required to produce a priming effect should be the category membership of the prime and target. The results of Experiment 2 are consistent with these predictions. First, there was a significant overall effect of prime in Experiment 2, replicating the findings of prior research (e.g., Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), although this effect was smaller in the present research than what has been produced in prior studies. Second, the largest predictors from among the subject-specific proficiency measures were the percentage of time that subjects used English at school, their self-rated English speaking proficiency, and the frequency of English use in social settings outside of home and school, while self-rated reading, speaking, and writing abilities in Chinese were all negative predictors. Third, the largest item-specific predictor of priming was no longer the frequency of the prime, but the frequency of the target, suggesting that priming in semantic categorization requires the exemplars to be ones which the subjects are exposed to frequently. This finding stands in direct contrast to the results in lexical decision, where target frequency had a negative impact on priming effects, once again consistent with the notion that some of the processes that drive priming in semantic categorization and lexical decision are qualitatively different. While still a positive predictor, the effect of prime frequency on the priming effect was considerably weaker than it was in lexical decision.

One issue with these types of analyses, of course, is that the criterion measures (in this case sPIP, iPIP, and PIP) used to derive predictions about the priming effects of subjects and items (based on subject LEQ responses, and prime length, frequency, and target frequency and stroke count) was fit using the same subjects and items used in the analysis. It is possible, then, that while the coefficients were well specified to make predictions on the data in Experiments 1 and 2, that the coefficients would not successfully predict priming effects on a new set of data that was not used to fit the predictive models. Finding evidence that these patterns would replicate with another dataset, then, would provide more compelling evidence that the model is not simply a model for the data already collected. To address this concern, Experiment 3 was a direct replication of

Experiment 2 and the model fitting was done using the parameters derived in the Experiment 2 analysis.

Chapter 4

4 Experiment 3

4.1 Method

4.1.1 Subjects

Subjects were 31 students (24 female, 7 male) at the University of Western Ontario, who participated in Experiment 3 for course credit. Subjects ranged between 18 to 29 years of age (*M* = 20.29, *SD* = 2.50). Of these subjects, 30 were right-handed, and one was left-handed. All subjects had normal or corrected-to-normal vision. Twenty-seven of the participants reported speaking Mandarin and English, while four participants reported being trilingual. One participant reported speaking Mandarin, English, and Japanese, two participants reported speaking Cantonese, Mandarin, and English, and one participant reported speaking Mandarin, English, and Spanish. Three participants could thus read in additional orthographic systems, as the Cantonese-Mandarin-English trilinguals could read in both traditional and simplified Chinese script, and the Mandarin-English-Japanese trilinguals could read in Japanese kana and Kanji.

4.1.2 Stimuli

The stimuli used in Experiment 3 were identical to the stimuli that were used in Experiment 2.

4.1.3 Procedure

The procedure was identical to that in Experiment 2, except instead of Experiment 3 being accompanied by a lexical decision task, Experiment 3 was accompanied by a speeded episodic recognition task (see Experiment 4 below). As with Experiments 1 and 2, the order in which Experiments 3 and 4 were performed by subjects was counterbalanced.

4.2 Results

4.2.1 Data Trimming

Data were trimmed using the same screening procedure as that used in Experiments 1 and 2. However, because the goal of Experiments 3 and 4 was not to directly compare their results by treating experiment as a fixed effect, the data were trimmed for Experiment 3 and 4 separately. During the first stage of the screening procedure, three items (1.5% of the total data) were discarded due to having error rates above 50%. During the second stage, six subjects (15.89% of the total data), and eight items (3.35% of the total data) were discarded due to being significant outliers in speed-accuracy space. Afterwards, the errors were separated from the correct responses (3.84% of the total data), and response times that deviated by more than three standard deviations from each subject's mean, or were less than 250 ms or greater than 2000 ms were discarded (2.53% of the total data; see Van Selst & Jolicouer, 1994), leaving approximately 73% of the data for analysis[18].

4.2.2 PIP

The coefficients for the PIP scores for Experiment 3 were the same as those used in Experiment 2, and can be found in Table 7. Additionally, alternative PIP scores were computed on the Experiment 3 data and those scores were also used in a second analysis. Finally, in a third analysis, Experiment 3's data were also analyzed together with Experiment 2's data, initially using the PIP coefficients derived from Experiment 2, and then the PIP coefficients derived from the combination of Experiment 2's and Experiment 3's data. The PIP coefficients derived from Experiment 3 can be found in Table 8, the sPIP, iPIP, and the means and standard deviations for the sPIP, iPIP and PIP coefficients are shown in Table 9. The Experiment 3 PIP coefficients indicated that the largest subject-based predictors were the usage of English at school and in other social contexts, English reading and speaking proficiency, and Chinese listening and writing proficiency. Negative predictors included English listening proficiency, and Chinese

[18] A follow-up analysis with loosened screening criteria retained roughly 82% of the overall data. The results of these analyses are reported in footnotes when they differ from the reported results.

reading proficiency. Target frequency and prime length were positive item-based predictors, while prime frequency and stroke count were negative predictors.

Table 8. *PIP coefficients for Experiment 3, Experiment 3 data only and combined data from Experiments 2 and 3.*

Predictor	Experiment 3 Coefficients	Combined Coefficients
sPIP		
PES	18.29	2.38
CW	1.07	0.00
PEO	11.67	6.78
CL	3.38	1.57
EL	-3.31	0.21
CR	-7.97	-3.52
ES	4.36	1.13
ER	9.62	0.00
iPIP		
GF	12.89	7.18
L	0.62	2.47
PCEL	-1.08	-2.72
NS	-6.72	-5.12

Note: PES = Percentage of time English is used at school; CW = Self-rated Chinese writing proficiency; PEO = Percentage of time English is used in other social settings; CL = Self-rated Chinese listening proficiency; EL = Self-rated English listening proficiency; CR = Self-rated Chinese reading proficiency; ES = Self-rated English speaking proficiency; ER = Self-rated English reading proficiency; GF = Target Google frequency; L = Prime length; PCEL = Prime CELEX frequency; NS = Target stroke count.

Table 9. *Mean sPIP, iPIP, and PIP Scores for Experiment 3.*

Measure	*M*	*SD*
sPIP	94.79	54.87
iPIP	-433.16	879.58
PIP	0.00	1.00

4.2.3 Reaction Time Analysis, Experiment 2 Coefficients

4.2.3.1 Prime x sPIP Analysis

As with Experiment 2, all analyses in Experiment 3 were conducted on the exemplar data. For the nonexemplar data, the results are described and shown in Appendix F. The Bayes Factor favoured the additive model over the full model, $BF = 2.73$, and a restricted model that excluded the main effect of sPIP, $BF = 2.72$, while the relative likelihood favoured both the full model, $\theta = 6.24$, and the restricted model, $\theta = 6.25$.

The models that included the interaction involved a main effect of prime, $\beta = -17.93$, $SE = 6.85$, $t(2141) = -2.62$, $p = .009$. Overall, targets that were preceded by a translation prime ($M = 641$ ms) produced faster latencies than targets that were preceded by a control prime ($M = 650$ ms), although this priming effect was rather small. The model also involved a null effect of sPIP, $\beta = 28.14$, $SE = 19.19$, $t(2141) = 1.47$, $p = .14$[19], and a two-way interaction between prime and sPIP, $\beta = -20.61$, $SE = 8.31$, $t(2141) = -2.48$, $p = .013$[20]. The two-way interaction between prime and sPIP is shown in Figure 13. As Figure 13 shows, priming effects were again larger for subjects who reported using English more in school and in other social contexts, reported having higher speaking and listening proficiency in English, and higher listening proficiency, but lower reading, speaking, and writing proficiency in Chinese. Response times increased as sPIP increased when the targets were preceded by a control prime, but stayed the same when the targets were preceded by a translation prime.

[19] The effect of sPIP was significant when the screening criteria were loosened, $t(2443) = 3.91$, $p < .0001$.
[20] The two-way interaction was nonsignificant when the screening criteria were loosened, $t(2443) = -1.50$, $p = .13$.

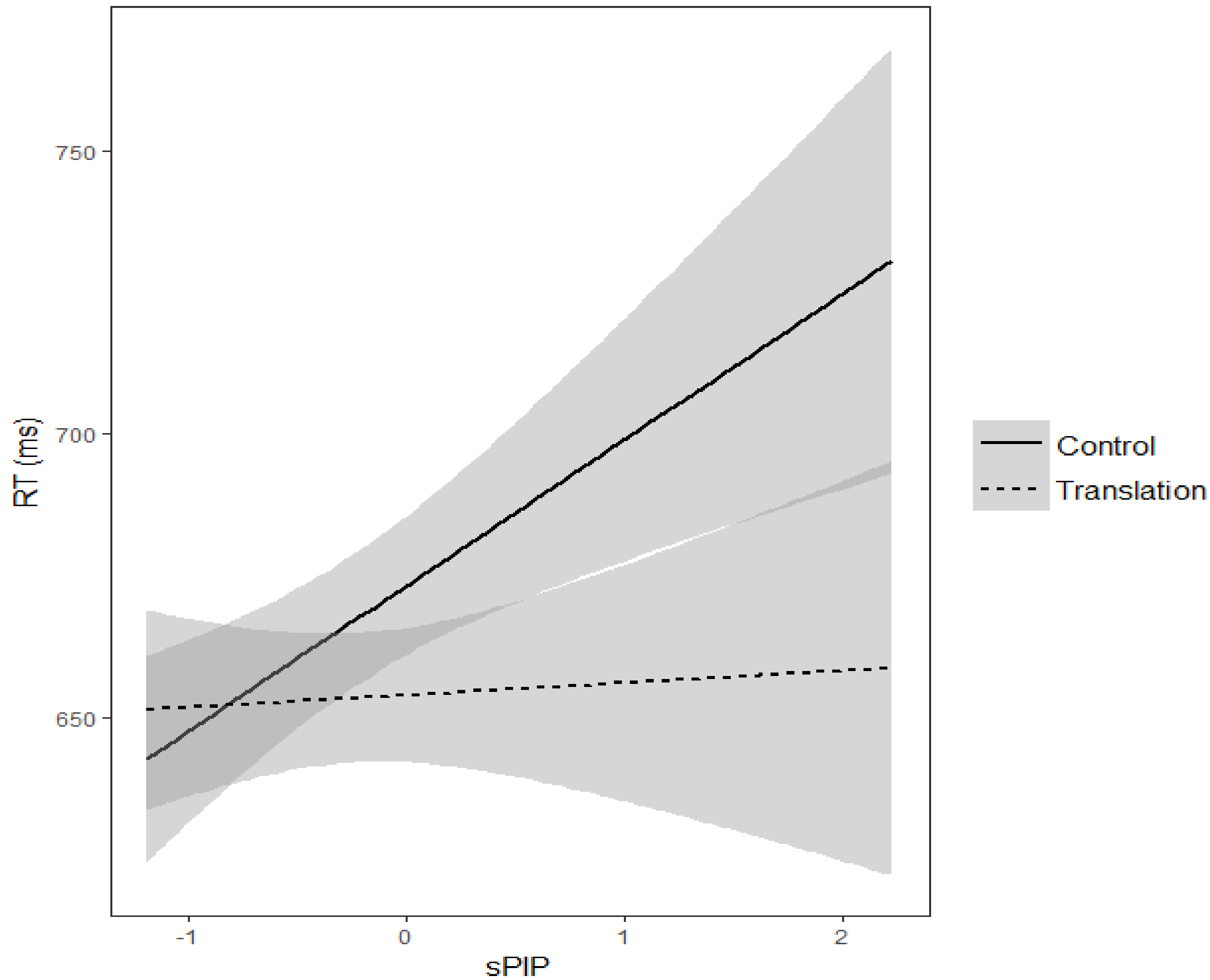

Figure 13. Response times as a function of prime and scaled sPIP, Experiment 3 exemplars, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

4.2.3.2 Prime x iPIP Analysis

The Bayes Factor favoured a restricted model which excluded the main effect of iPIP from the analysis, but retained the effects of prime and the two-way interaction over the fully interactive model, $BF = 7.23$, $\theta = 0.42$, and the additive model, $BF = 1.19$, $\theta = 1.19$. This analysis involved main a main effect of prime, $\beta = -12.66$, $SE = 6.35$, $t(2141) = -2.00$, $p = .046$, a null effect of iPIP, $t < 1$, and a significant two-way interaction between prime and iPIP, $\beta = -15.30$, $SE = 7.27$, $t(2141) = -2.10$, $p = .035$. The two-way interaction between prime and iPIP is shown in Figure 14. As shown in Figure 14, priming effects were again larger when Chinese targets were higher in frequency, and when the prime was longer than when the targets were low-frequency, and short primes were used. The effect of iPIP on RTs varied as a function of prime type. Response times to targets preceded by translation primes decreased as iPIP increased, while RTs for targets preceded by control primes stayed the same.

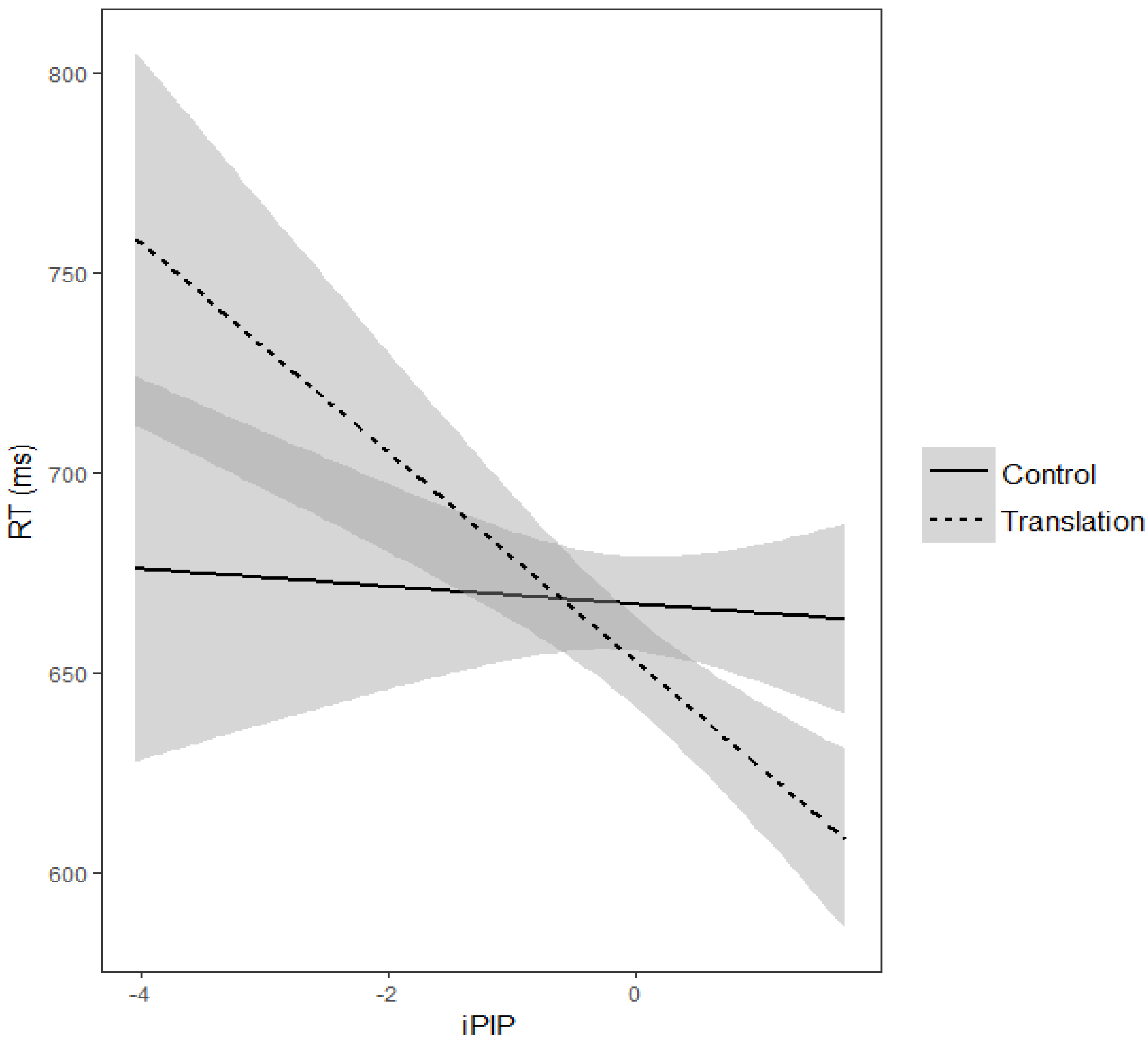

Figure 14. Response times as a function of prime and scaled iPIP, Experiment 3 exemplars, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

4.2.3.3 Prime x PIP Analysis

The fully interactive model was favoured over the additive model, *BF* = 1.14, θ = 19.39, and involved a significant effect of prime, β = -18.42, *SE* = 6.88, $t(2141) = -2.68$, $p = .007$, and a significant two-way interaction between prime and PIP, $\beta = -23.10$, $SE = 8.04$, $t(2141) = -2.87$, $p = .0041$. The effect of PIP was nonsignificant, $t < 1$. The two-way interaction between prime and PIP is shown in Figure 15. At lower PIP scores, an inhibitory effect of prime occurs, while at higher PIP scores, a priming effect is produced. In sum, the combination of subject- and item-specific factors used to compute the sPIP and iPIP scores, as derived from subjects in Experiment 2, predicted priming effects in Experiment 3.

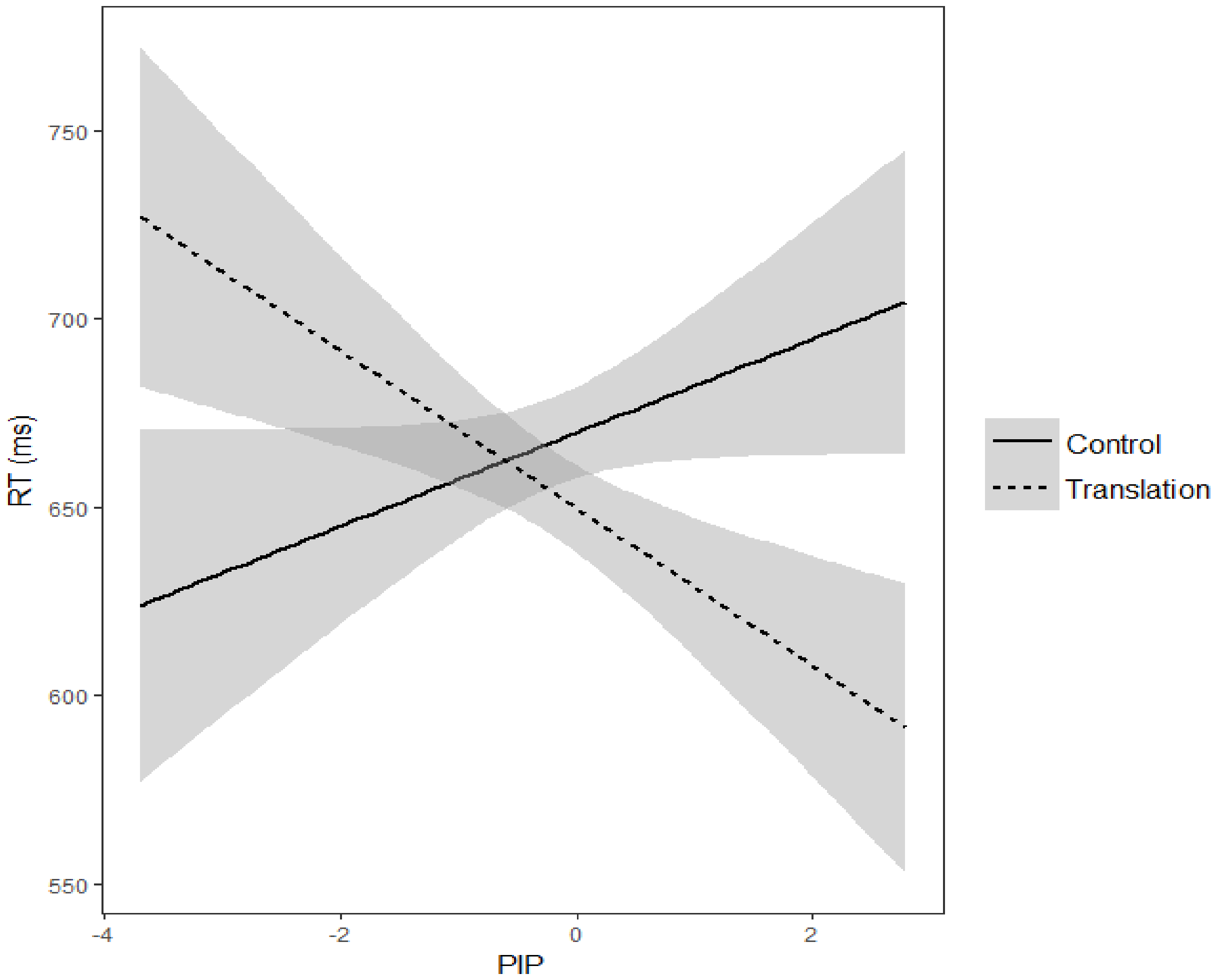

Figure 15. Response times as a function of prime and PIP score, Experiment 3 exemplars, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

4.2.4 Error Analysis, Experiment 2 Coefficients

4.2.4.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, *BF* = 45.32, θ = 2.55. None of the effects were significant in any of the analyses, *z*s < 1.29, *p*s > .19[21]. The mean error rates as a function of prime and sPIP tertile are shown in Figure 16.

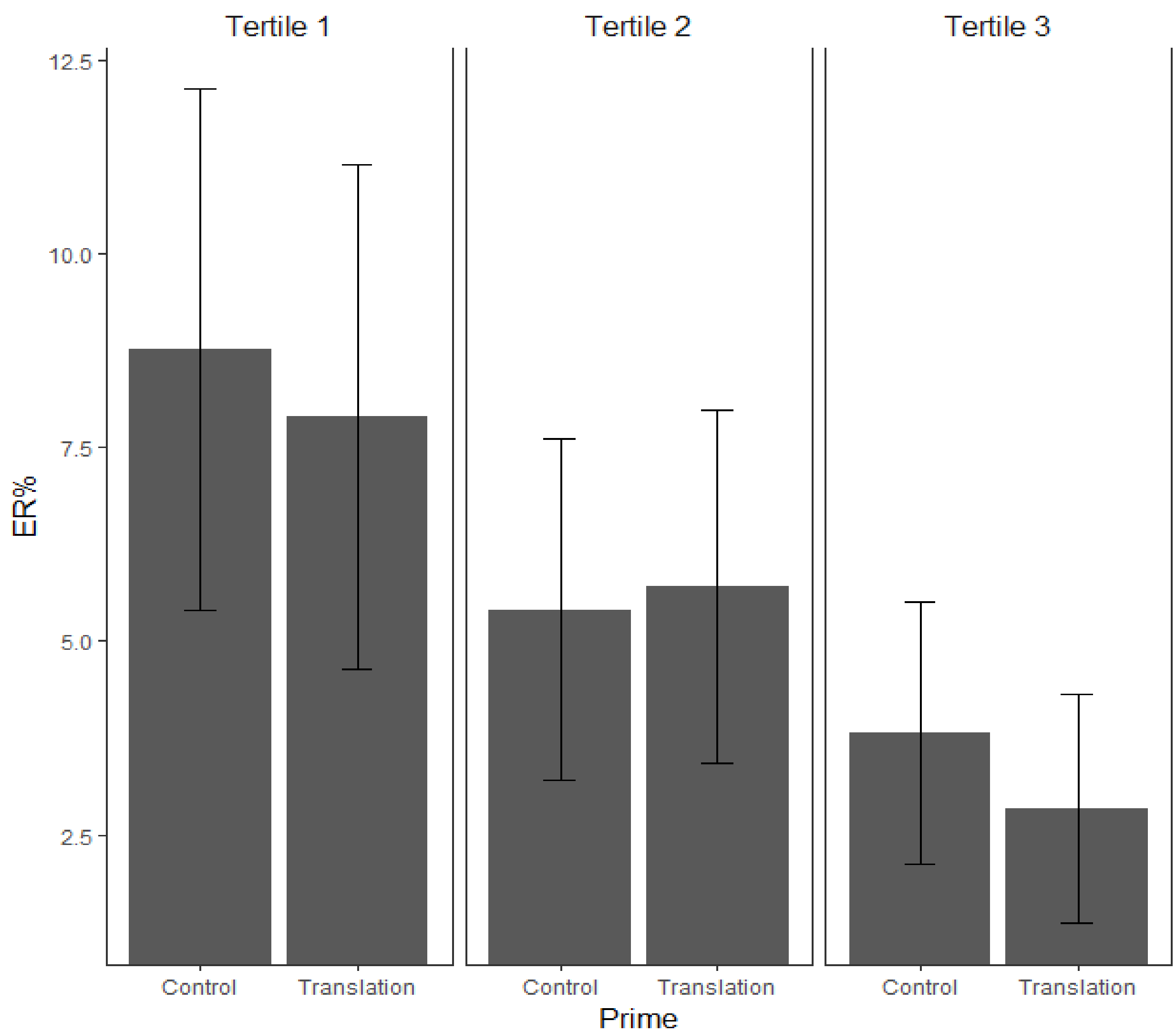

Figure 16. Mean error rates as a function of prime and sPIP tertile, Experiment 3 exemplars, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

[21] The effect of sPIP was significant when the screening criteria were loosened, *z*(2668) = -2.48, *p* = .013.

4.2.4.2 Prime x iPIP Analysis

The additive model was favoured over the interactive model, *BF* = 45.95, θ = 2.58. None of the effects were significant in any of the analyses, *zs* < 1. The mean error rates as a function of prime and iPIP tertile are shown in Figure 17.

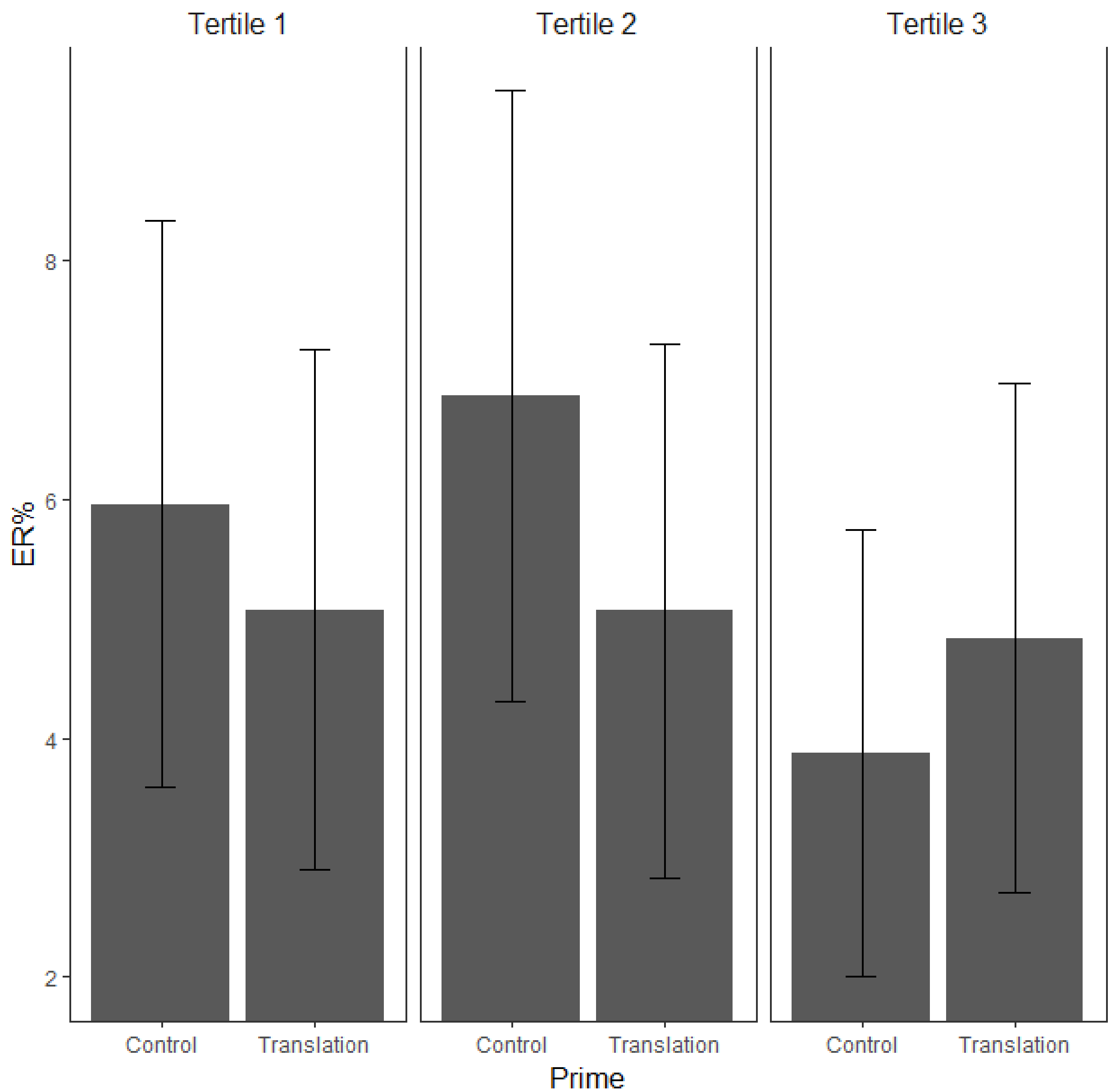

Figure 17. Mean error rates as a function of prime and iPIP tertile, Experiment 3 exemplars, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

4.2.4.3 Prime x PIP Analysis

The additive model was again favoured over the interactive model, *BF* = 48.37, θ = 2.72. Once again, none of the effects were significant in any of the analyses, *z*s < 1. The mean error rates as a function of prime and PIP tertile are shown in Figure 18.

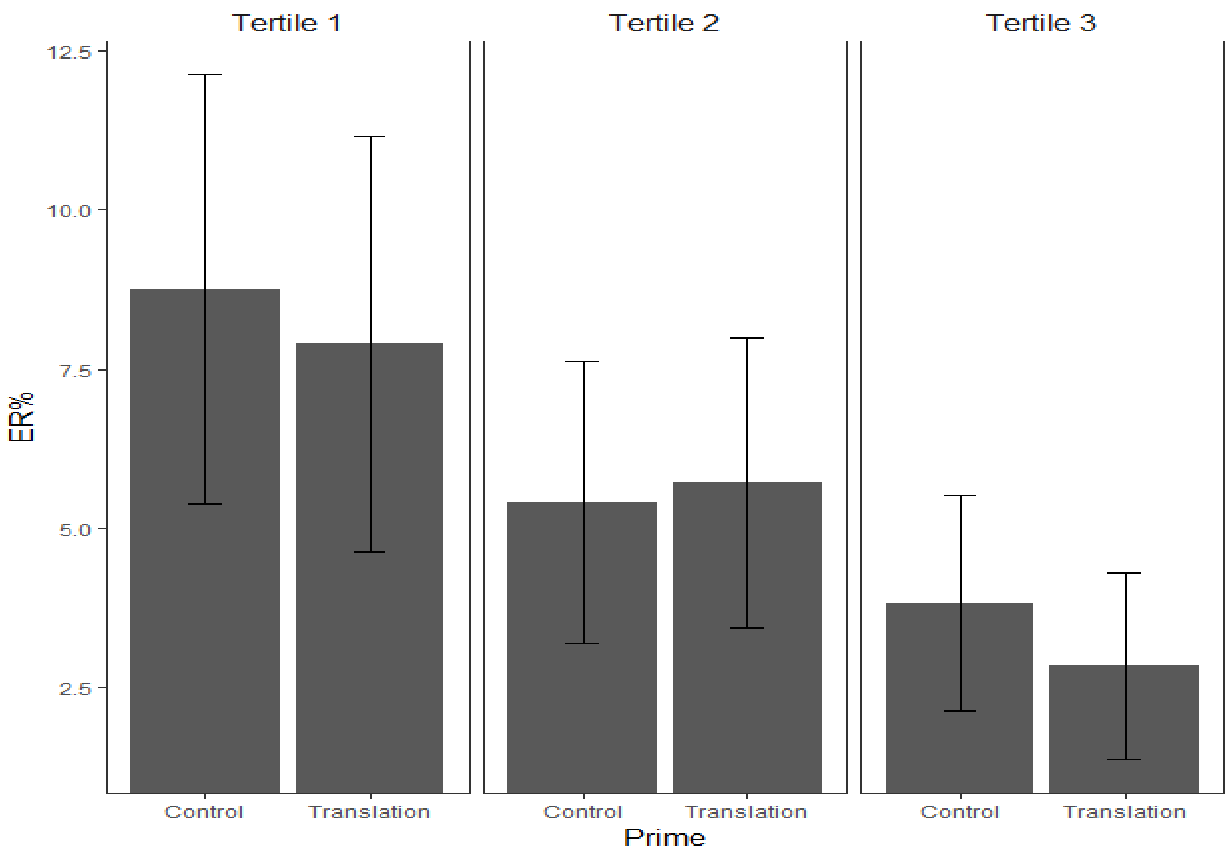

Figure 18. Mean error rates as a function of prime and PIP tertile, Experiment 3 exemplars, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

4.2.5 Reaction Time Analysis, Experiment 3 Coefficients

4.2.5.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 1.84, but not in the relative likelihood analysis, θ = 0.11. In the interactive model, the main effect of prime, β = -13.01, *SE* = 6.27, *t*(2147) = -2.08, *p* = .038, and the two-way interaction between prime and sPIP were significant, β = -17.26, *SE* = 6.68, $t(2147) = -2.58$, $p = .01$[22], while the main effect of sPIP was nonsignificant, β = 19.09, *SE* = 17.38, $t(2147) = 1.10$, $p = .27$[23]. The two-way interaction is shown in Figure 19. Priming effects were larger for subjects who reported using English more at school and in other social contexts, reported higher reading and speaking proficiency, but lower listening proficiency in English, and higher listening and writing proficiency, but lower reading proficiency in Chinese.

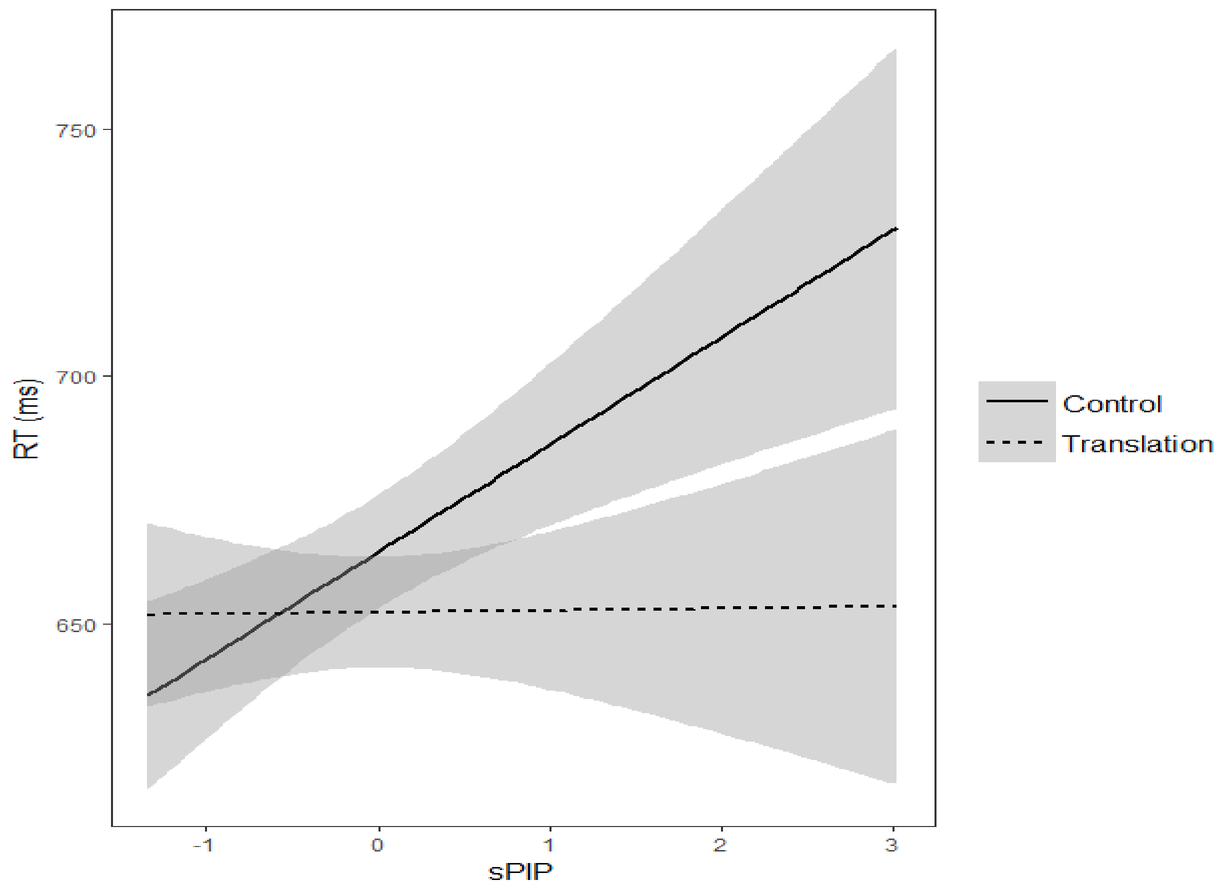

Figure 19. Response times as a function of prime and scaled sPIP, Experiment 3 exemplars, Experiment 3 coefficients. Shaded areas represent 95% confidence intervals.

[22] The two-way interaction was marginally significant when the screening criteria were loosened, $t(2401) = -1.69$, $p = .09$.
[23] The effect of sPIP was significant when the screening criteria were loosened, $t(2401) = 3.83$, $p = .0001$.

4.2.5.2 Prime x iPIP Analysis

The additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 1.56, but not in the relative likelihood analysis, θ = 0.09. In the fully interactive model, the main effect of prime, *β* = -14.08, *SE* = 6.39, *t*(2147) = -2.20, *p* = .028, and the two-way interaction between prime and iPIP were significant, *β* = -19.21, *SE* = 7.30, *t*(2147) = -2.63, *p* = .0085, while the effect of iPIP was nonsignificant, *t* < 1. The two-way interaction is shown in Figure 20. Priming effects were larger when targets were higher frequency, had fewer strokes, and when primes were longer in length and lower in frequency.

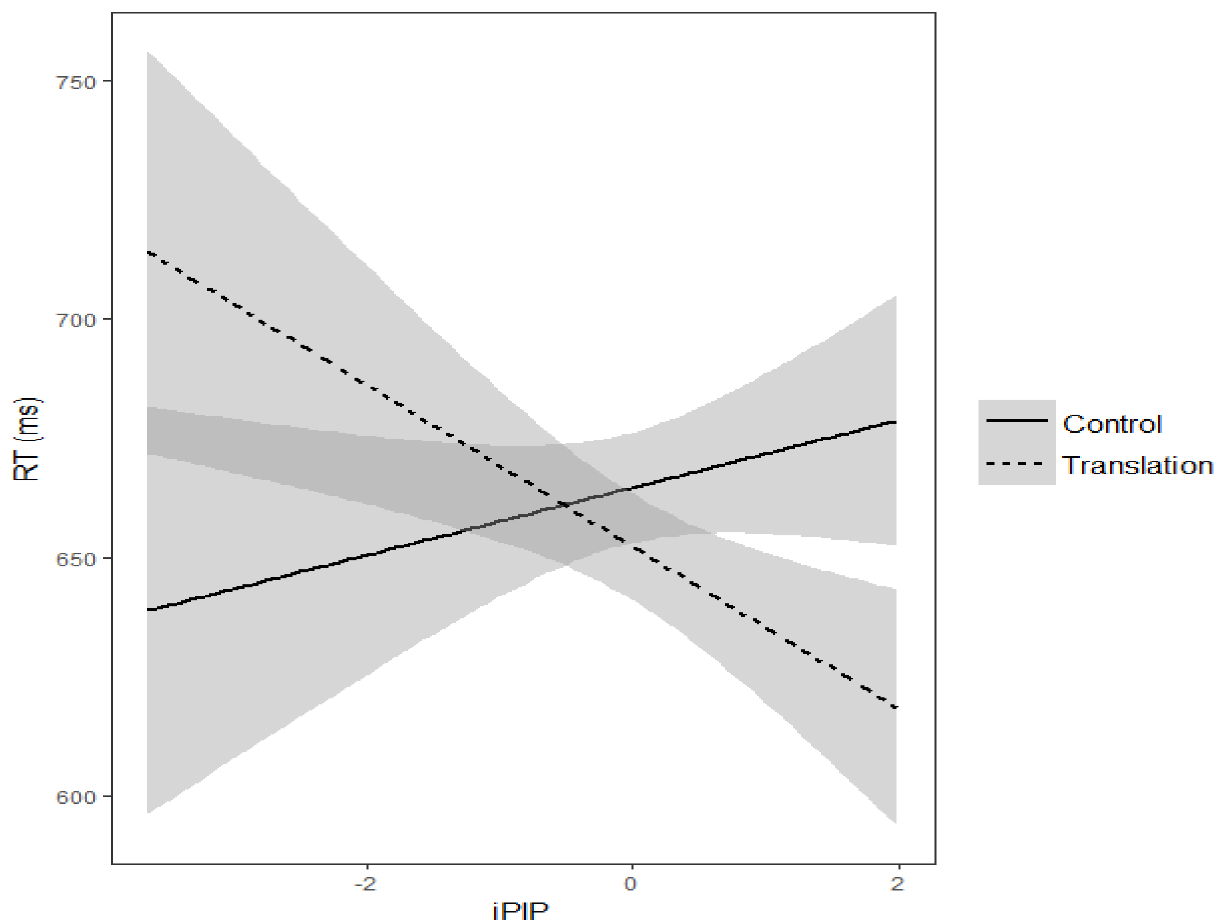

Figure 20. Response times as a function of prime and scaled iPIP, Experiment 3 exemplars, Experiment 3 coefficients. Shaded areas represent 95% confidence intervals.

4.2.5.3 Prime x PIP Analysis

The interactive model was favoured over the additive model, *BF* = 10.58, θ = 180.30. The main effect of prime, β = -15.51, *SE* = 6.54, *t*(2147) = -2.37, *p* = .018, and the two-way interaction between prime and PIP were significant, β = -25.38, *SE* = 7.00, *t*(2147) = -3.62, *p* = .0003, while the main effect of PIP was nonsignificant, β = 11.26, *SE* = 9.46, *t*(2147) = 1.19, *p* = .23[24]. The two-way interaction is shown in Figure 21. Overall, the combination of subject- and item-specific factors that were used to compute the sPIP and iPIP scores, as derived from subjects in Experiment 3, predicted priming effects in Experiment 3.

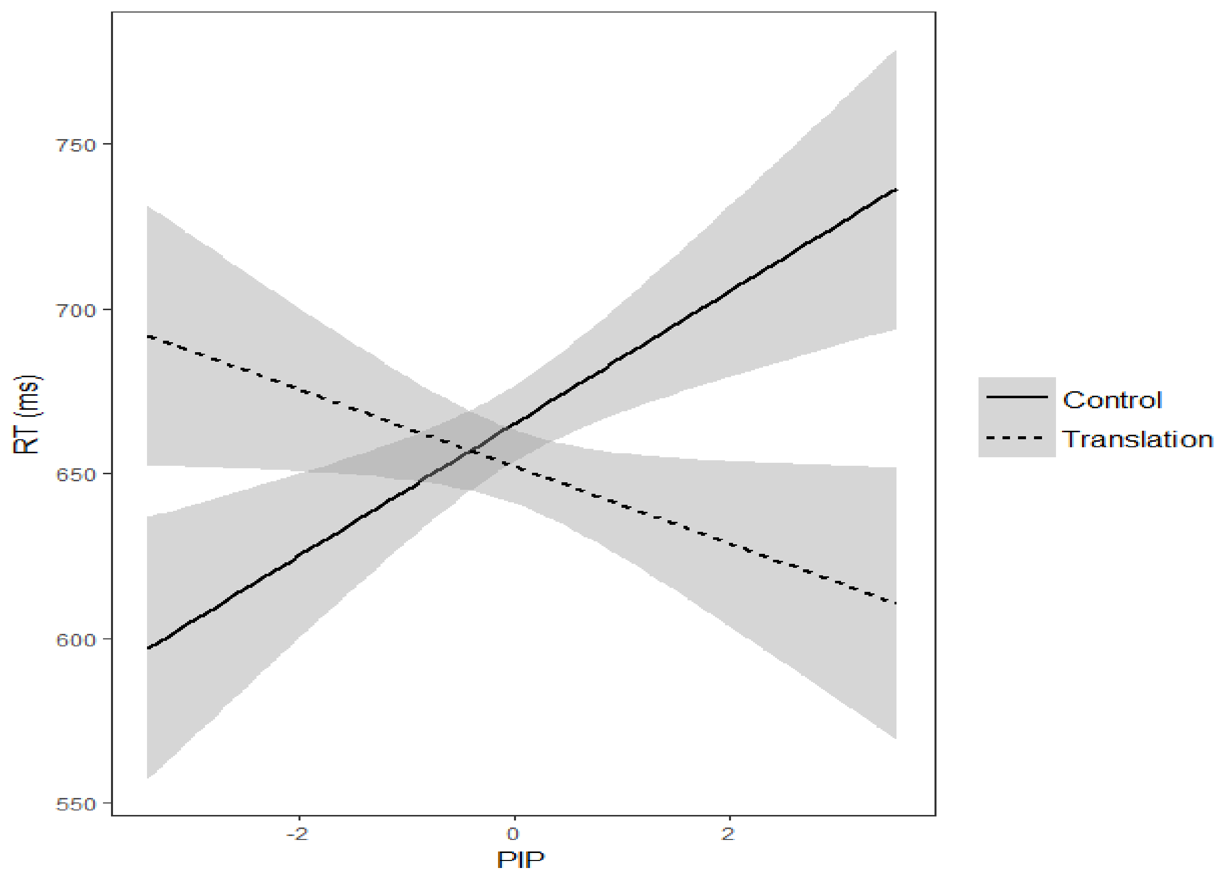

Figure 21. Response times as a function of prime and PIP, Experiment 3 exemplars, Experiment 3 coefficients. Shaded areas represent 95% confidence intervals.

[24] The effect of PIP was significant when the screening criteria were loosened, *t*(2401) = 2.36, *p* = .018.

4.2.5.4 Prime x Experiment Order Analysis

The only significant effect in this analysis was the effect of prime, β = -39.46, *SE* = 17.01, $t(2147) = -2.32$, $p = .01$. Neither the effect of order, nor the two-way interaction was significant, *t*s < 1.

4.2.5.4 Prime x List Analysis

The effect of prime was once again significant in this analysis, β = 18.77, *SE* = 8.65, $t(2147) = -2.17$, $p = .015$. In addition, the effect of list was significant, β = 71.46, *SE* = 20.27, $t(2147) = 3.52$, $p = .0004$. Response latencies in List 1 (*M* = 619 ms) were significantly faster than latencies in List 2 (*M* = 672 ms). Most importantly, the two-way interaction between prime and list was nonsignificant, β = -18.86, *SE* = 14.66, $t(2147) = -1.29$, $p = .20$. There was no significant difference in the priming effect in List 1 (8 ms) and List 2 (10 ms).

4.2.6 Error Analysis, Experiment 3 Coefficients

4.2.6.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, *BF* = 46.80, θ = 2.63. None of the effects were significant in any analysis, *z*s < 1.28, *p*s > .19[25]. The mean error rates as a function of prime and sPIP tertile are shown in Figure 22.

[25] The effect of sPIP was significant when the screening criteria were loosened, $z(2668) = -2.45$, $p = .014$.

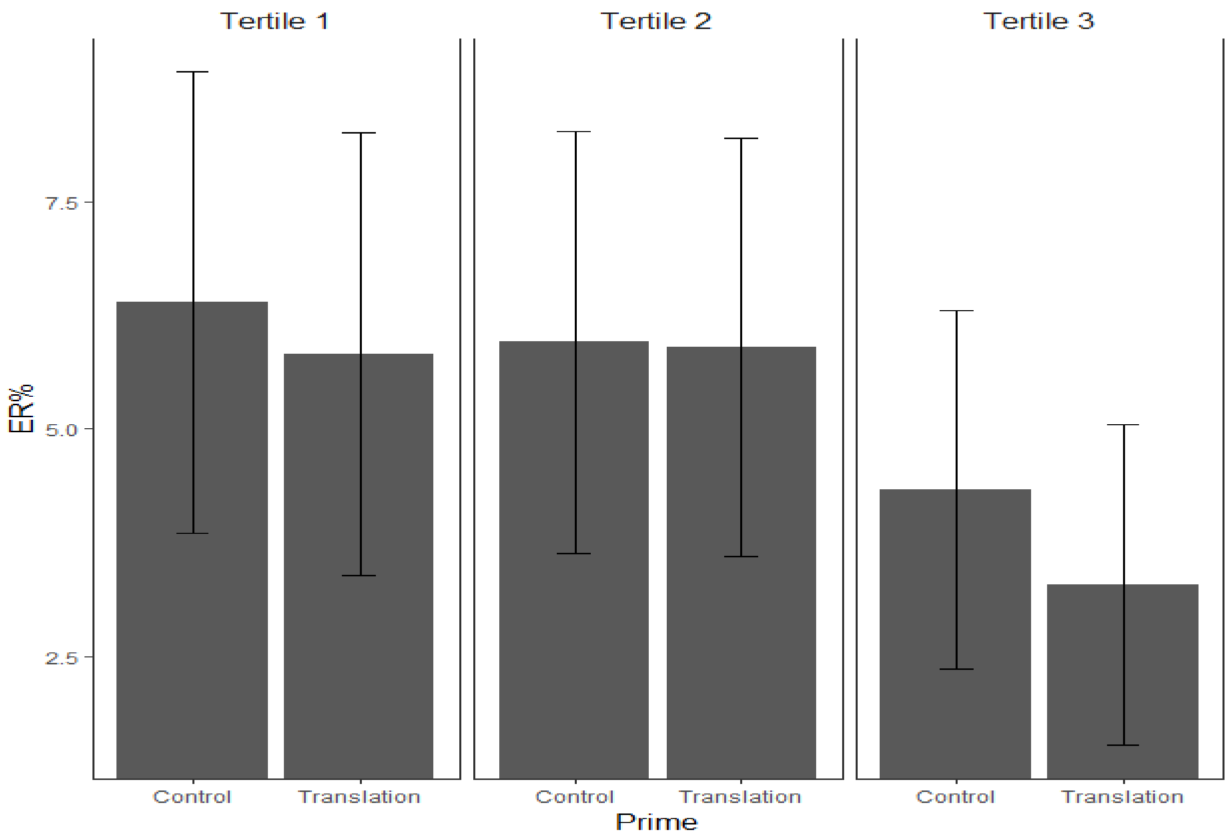

Figure 22. Mean error rates as a function of prime and sPIP tertile, Experiment 3 exemplars, Experiment 3 coefficients. Error bars represent 95% confidence intervals.

4.2.6.2 Prime x iPIP Analysis

The additive model was again favoured over the interactive model, *BF* = 47.69, θ = 2.68. None of the effects were significant in any analysis, *zs* < 1. The mean error rates as a function of prime and iPIP tertile are shown in Figure 23.

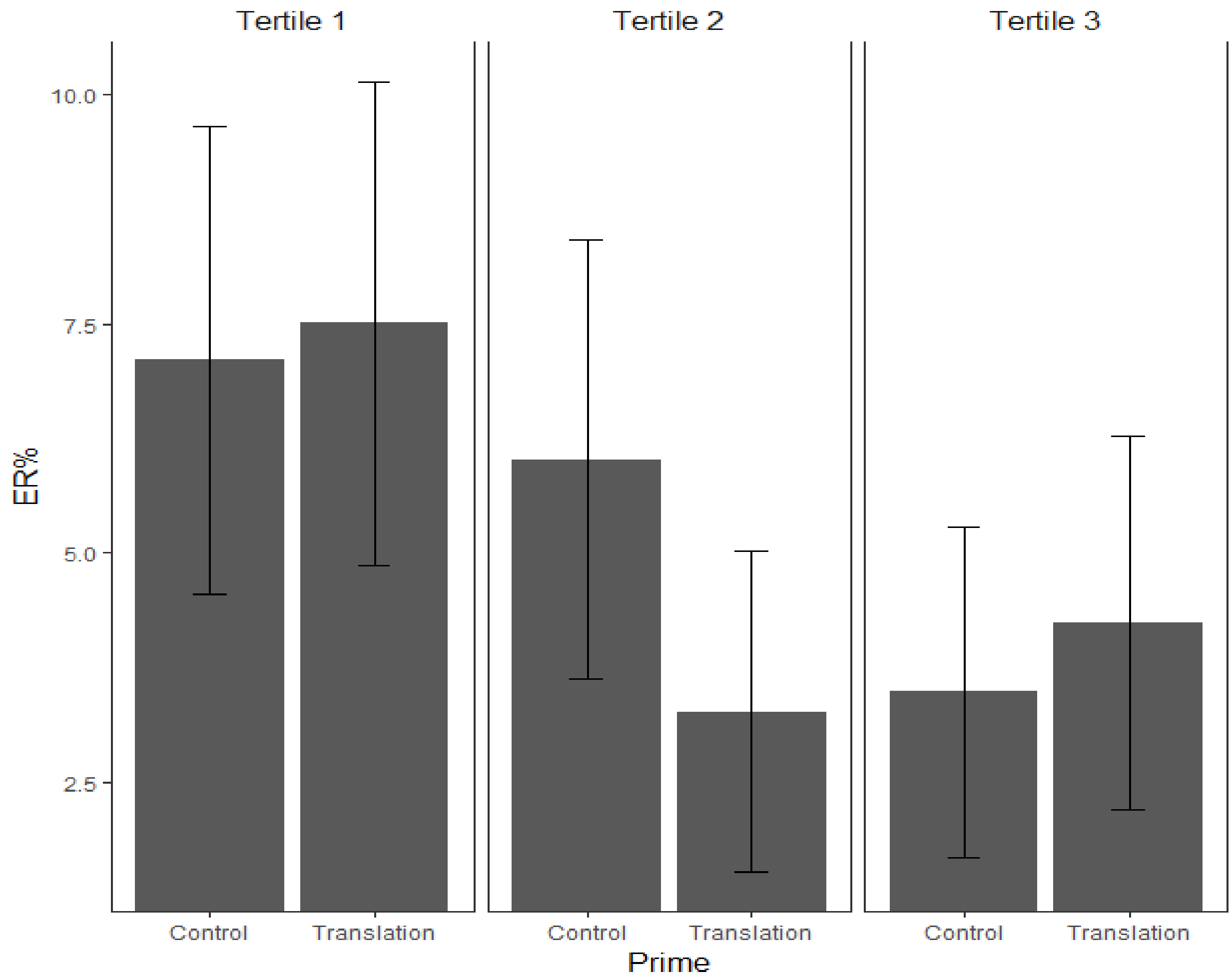

Figure 23. Mean error rates as a function of prime and iPIP tertile, Experiment 3 exemplars, Experiment 3 coefficients. Error bars represent 95% confidence intervals.

4.2.6.3 Prime x PIP Analysis

The additive model was again favoured over the interactive model, *BF* = 46.20, θ = 2.60. Again, none of the effects were significant in any analysis, *z*s < 1. The mean error rates as a function of prime and PIP tertile are shown in Figure 24.

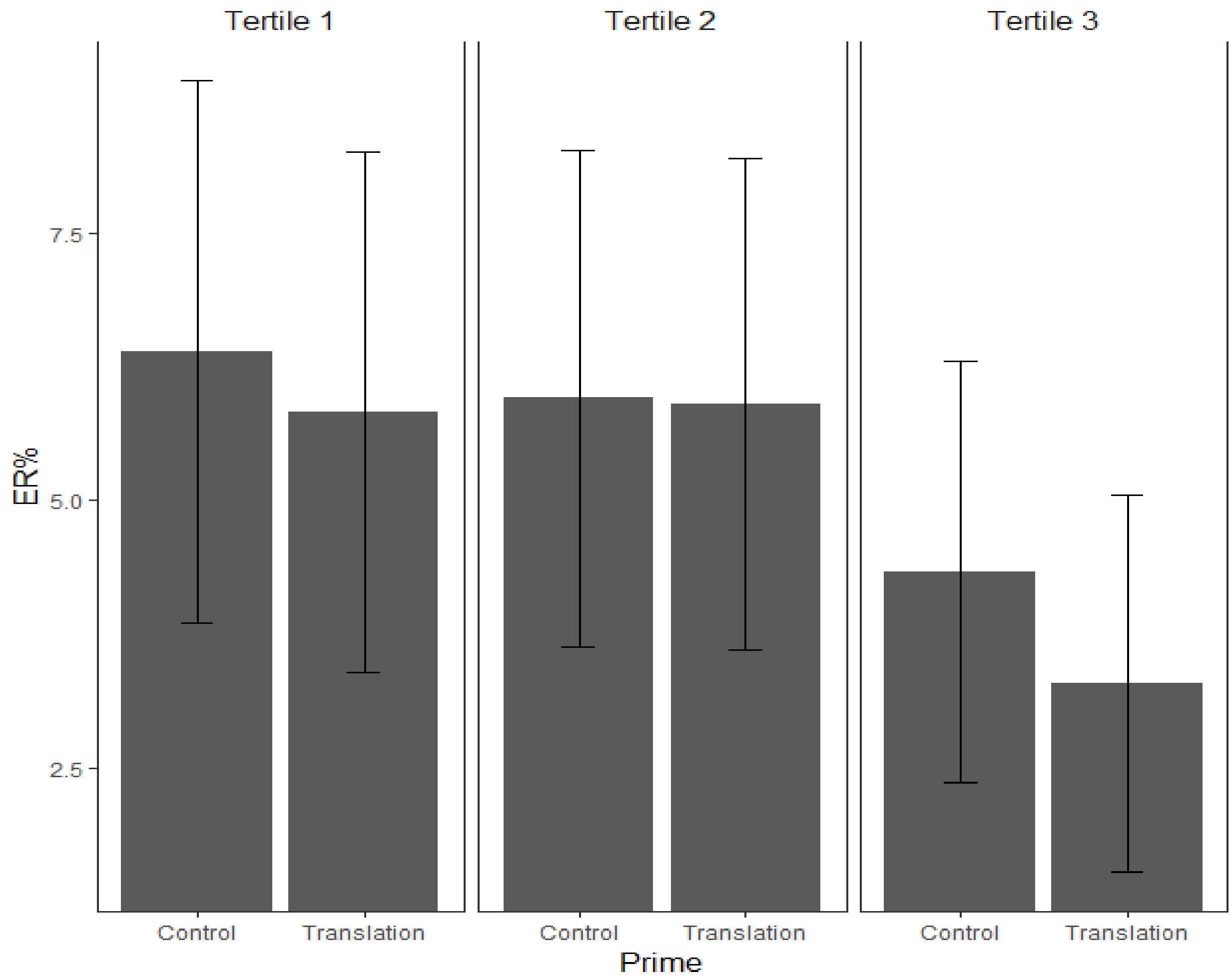

Figure 24. Mean error rates as a function of prime and PIP tertile, Experiment 3 exemplars, Experiment 3 coefficients. Error bars represent 95% confidence intervals.

Chapter 5

5 Combined Analysis of Experiments 2 and 3

Before moving on to Experiment 4, a final series of analyses were conducted on the combined data from Experiments 2 and 3 to assess how well the coefficients derived from Experiment 2 sufficiently account for the overall data from both experiments. One series used the coefficients from Experiment 2 whereas the other used the coefficients derived from the combined data of both experiments. Additionally, sPIP, iPIP and PIP scores were derived from the overall data to assess what factors best accounted for the priming data in the overall data.

5.1 Results

5.1.1 PIP

The PIP coefficients derived from the combined data are shown in Table 8. The iPIP coefficients derived from the combined data with typicality accounted for are found in Table 10. Additionally, the means and standard deviations for the sPIP, iPIP, and PIP scores derived from the combined data are shown in Table 11. In the overall coefficients, the largest subject-based predictors were the use of English in other social contexts, the use of English at school, Chinese listening proficiency, English speaking proficiency, and English listening proficiency. Chinese reading proficiency was the only negative predictor for the full data. Without typicality, the largest facilitative item-based predictors were target frequency and prime length, while the number of strokes and prime frequency were the largest negative predictors. With typicality, the largest item-based predictors were target frequency and category typicality. Both prime length and frequency had a facilitative influence on priming, but the effect was relatively weak. The only inhibitory factor was the number of target strokes.

Table 10. *iPIP coefficients for Experiment 3, typicality included, combined data from Experiments 2 and 3.*

Predictor	Experiment 3 Coefficients
GF	6.12
TYP	2.27
L	0.70
PCEL	0.63
NS	-1.40

Note: GF = Target Google frequency; TYP = Prime category typicality ratings; L = Prime length; PCEL = Prime CELEX frequency; NS = Target stroke count.

Table 11. *Means and Standard Deviations for sPIP, iPIP, and PIP Scores for the Combined Experiments 2 & 3 Data.*

Coefficients	*M*	*SD*
Experiment 2 Coefficients		
sPIP	70.18	36.28
iPIP	71.66	100.21
PIP	0.00	1.00
Combined Coefficients		
sPIP	30.97	22.88
iPIP	-148.05	352.84
PIP	0.00	1.00
Typicality Included		
iPIP	0.82	5.63
PIP	0.00	1.00

5.1.2 Reaction Time Analysis, Experiment 2 Coefficients

5.1.2.1 Prime x sPIP Analysis

As with Experiments 2 and 3, all analyses were conducted on the exemplar data. For the nonexemplar data, all analyses are described and shown in Appendix F. The interactive model was favoured over the additive model, *BF* = 1.77, θ = 63.15, and involved a significant main effect of prime, β = -10.43, *SE* = 3.05, $t(9444) = -3.42$, $p = .0006$. In the combined data, targets that were preceded by translation primes (*M* = 667 ms) produced faster latencies than targets that were preceded by control primes (*M* = 676 ms). While there was no significant effect of sPIP, $t < 1$[26], there was a significant two-way interaction between prime and sPIP, β = -10.39, *SE* = 3.12, $t(9444) = -3.32$, $p = .0009$. The two-way interaction is shown in Figure 25. As shown in Figure 25, priming effects were larger for subjects who reported using English more at school and in other social contexts, reported higher speaking and listening proficiency in English, and higher listening proficiency, but lower reading, writing, and speaking proficiency in Chinese.

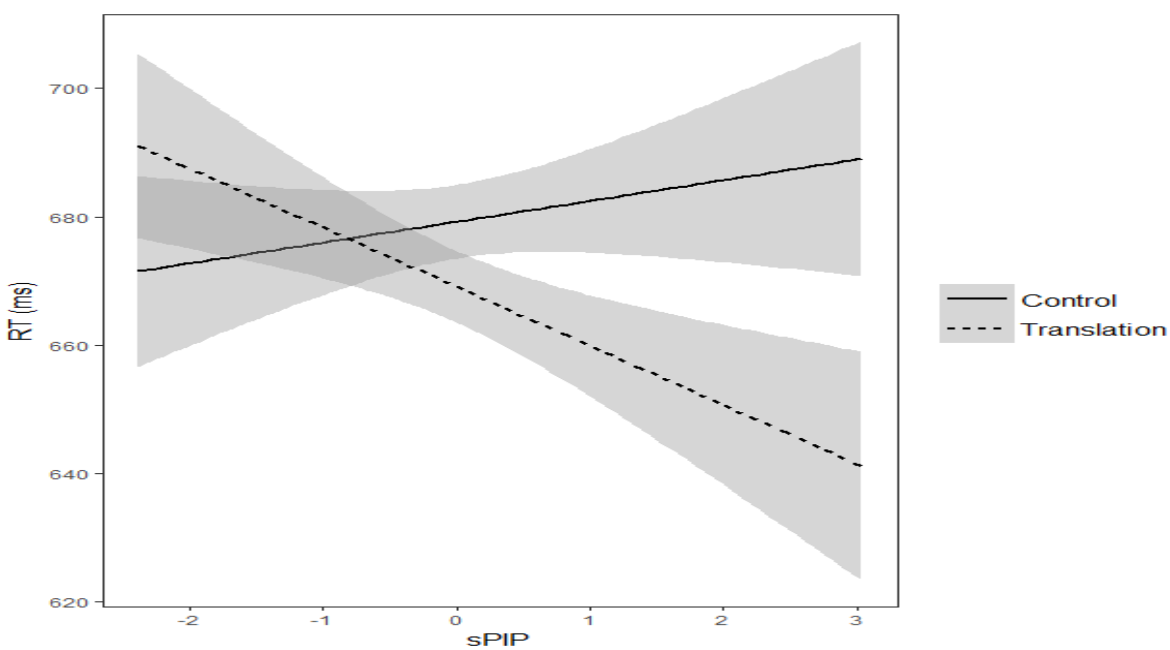

Figure 25. Response times as a function of prime and scaled sPIP, combined Experiments 2 and 3 exemplar data, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

[26] The effect of sPIP was marginally significant when the screening criteria were loosened, $t(10736) = 1.87$, $p = .06$.

5.1.2.2 Prime x iPIP Analysis

The interactive model with no random slopes was favoured over the interactive model that included iPIP as a random slope on items, *BF* = 347.64, θ = 0.27. The additive model with no slopes was favoured over this interactive model in the Bayes Factor analysis, *BF* = 6.06, but the interactive model was favoured in the relative likelihood analysis, θ = 5.90. While there was approximately six times greater likelihood that the data occurred under the assumptions of the additive model, there was almost an equal likelihood that excluding the interaction would result in significant data loss. The additive model involved a significant effect of prime, *β* = -10.27, *SE* = 2.96, *t*(9444) = -3.47, *p* = .0005, but a nonsignificant effect of iPIP, *β* = -2.93, *SE* = 2.67, *t*(9444) = -1.10, *p* = .27. Additionally, the interactive model involved a significant two-way interaction between prime and iPIP, *β* = -8.25, *SE* = 3.81, *t*(9444) = -2.17, *p* = .03, which is shown in Figure 26. As shown in Figure 26, priming effects were larger for high-frequency Chinese targets that were preceded by longer English primes.

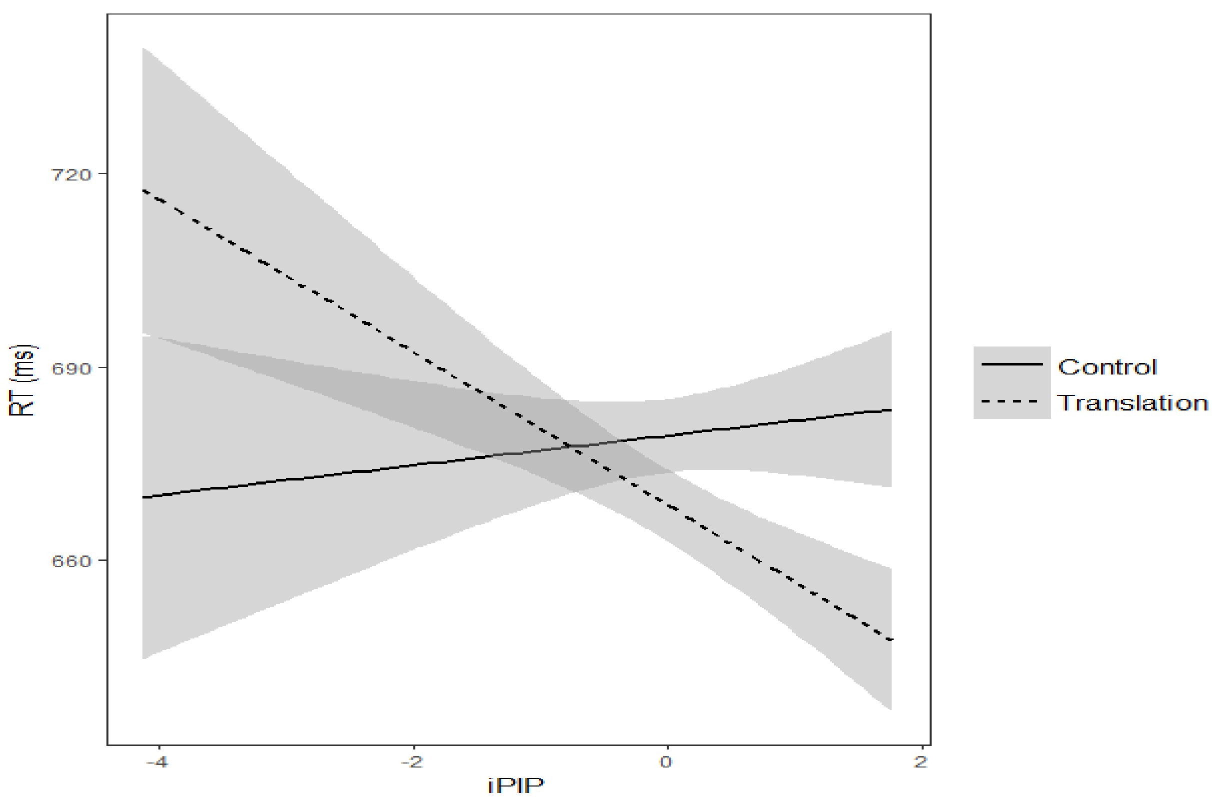

Figure 26. Response times as a function of prime and scaled iPIP, combined Experiment 2 and 3 exemplar data, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

5.1.2.3 Prime x PIP Analysis

The interactive model was favoured over the additive model, *BF* = 22.32, θ = 797.82, and involved a significant main effect of prime, β = -10.31, *SE* = 3.16, $t(9444) = -3.26$, $p = .0011$, and a significant two-way interaction between prime and PIP, β = -13.74, *SE* = 3.38, $t(9444) = -4.06$, $p < .0001$, but no effect of PIP, $t < 1$. As shown in Figure 27, the combination of subject- and item-specific factors that were used to compute the sPIP and iPIP scores, as derived from the Experiment 2 coefficients, significantly predicted priming effects in the combined data.

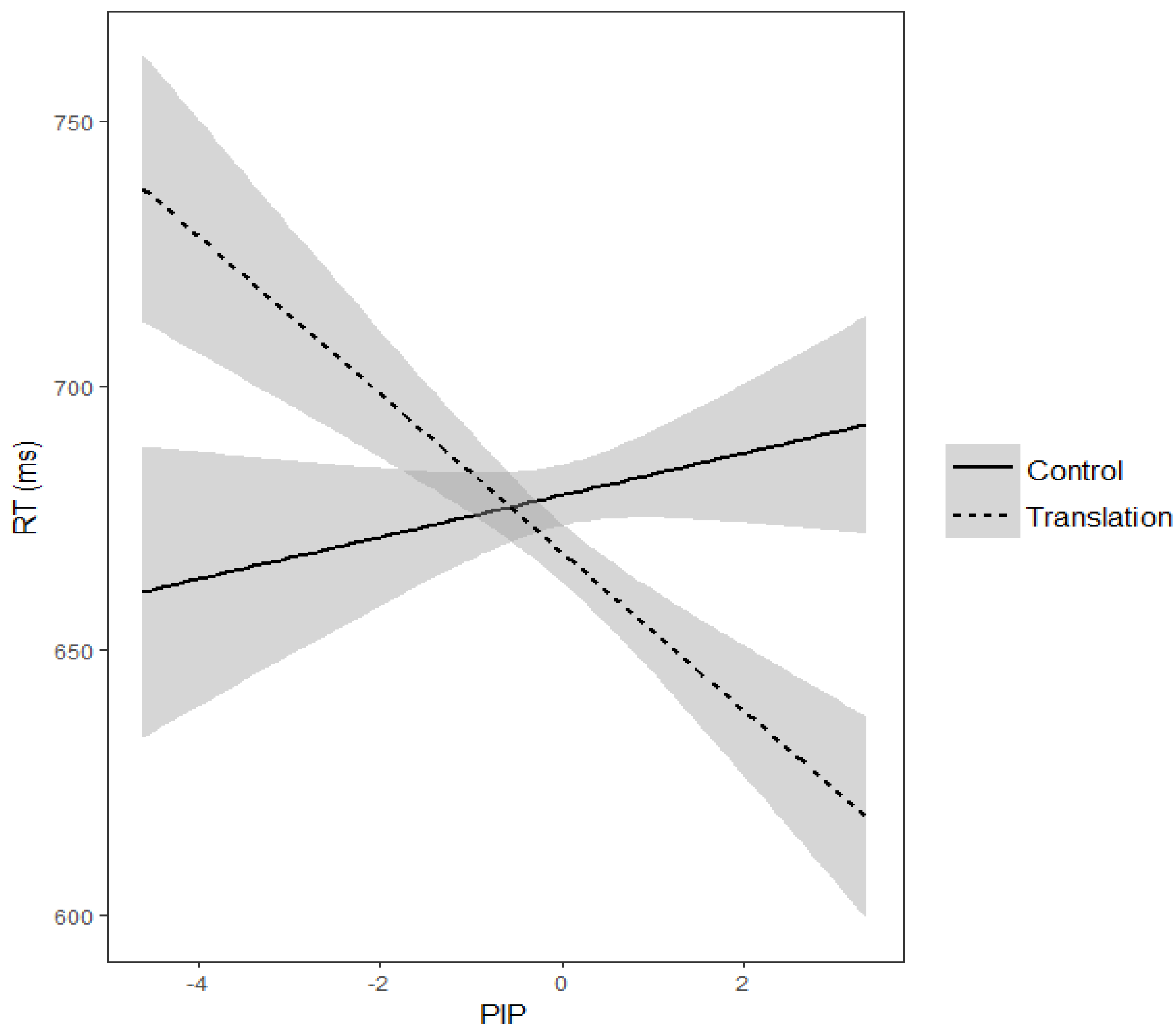

Figure 27. Response times as a function of prime and PIP, combined Experiment 2 and 3 exemplar data, Experiment 2 coefficients. Shaded areas represent 95% confidence intervals.

5.1.3 Reaction Time Analysis, Combined Coefficients

5.1.3.1 Prime x sPIP Analysis

The interactive model was favoured over the additive model, *BF* = 1.08, θ = 38.67, and involved a significant effect of prime, $\beta = -10.35$, *SE* = 2.94, *t*(9444) = -3.52, *p* = .0004, and a significant two-way interaction between prime and sPIP, $\beta = -9.60$, *SE* = 3.01, *t*(9444) = -3.19, *p* = .0014, while the effect of sPIP was nonsignificant, $t < 1$. The two-way interaction is shown in Figure 28. Priming effects were larger for subjects who reported using English more at school and in other social contexts, reported higher speaking and listening proficiency in English, and higher listening proficiency, but lower reading proficiency in Chinese.

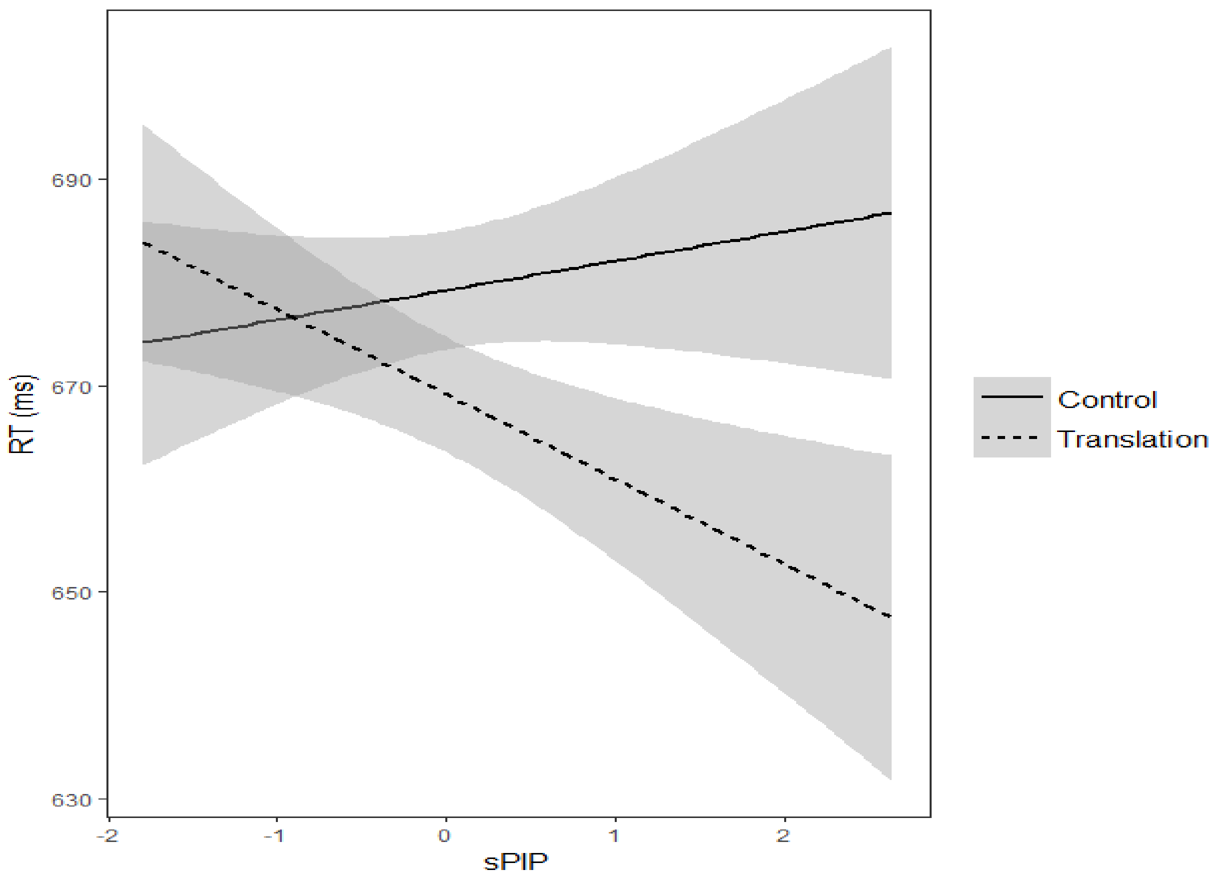

Figure 28. Response times as a function of prime and scaled sPIP, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Shaded areas represent 95% confidence intervals.

5.1.3.2 Prime x iPIP Analysis, Typicality Excluded

The additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 2.85, but the interactive model was favoured over the additive model in the relative likelihood analysis, θ = 12.56. Because the likelihood that excluding the two-way interaction between prime and iPIP would result in significant data loss was considerably larger than the difference in the amount of evidence consistent with each model, the interactive model was selected. This analysis involved a significant effect of prime, $\beta = -10.70$, $SE = 3.00$, $t(9444) = -3.57$, $p = .0004$, and a significant two-way interaction between prime and iPIP, $\beta = -9.80$, $SE = 3.54$, $t(9444) = -2.77$, $p = .0056$. The effect of iPIP was nonsignificant, $\beta = 3.18$, $SE = 2.38$, $t(9444) = 1.34$, $p = .18$[27]. The two-way interaction is shown in Figure 29. Priming effects were larger for higher frequency Chinese targets that had fewer strokes, and which were preceded by longer, lower-frequency English primes.

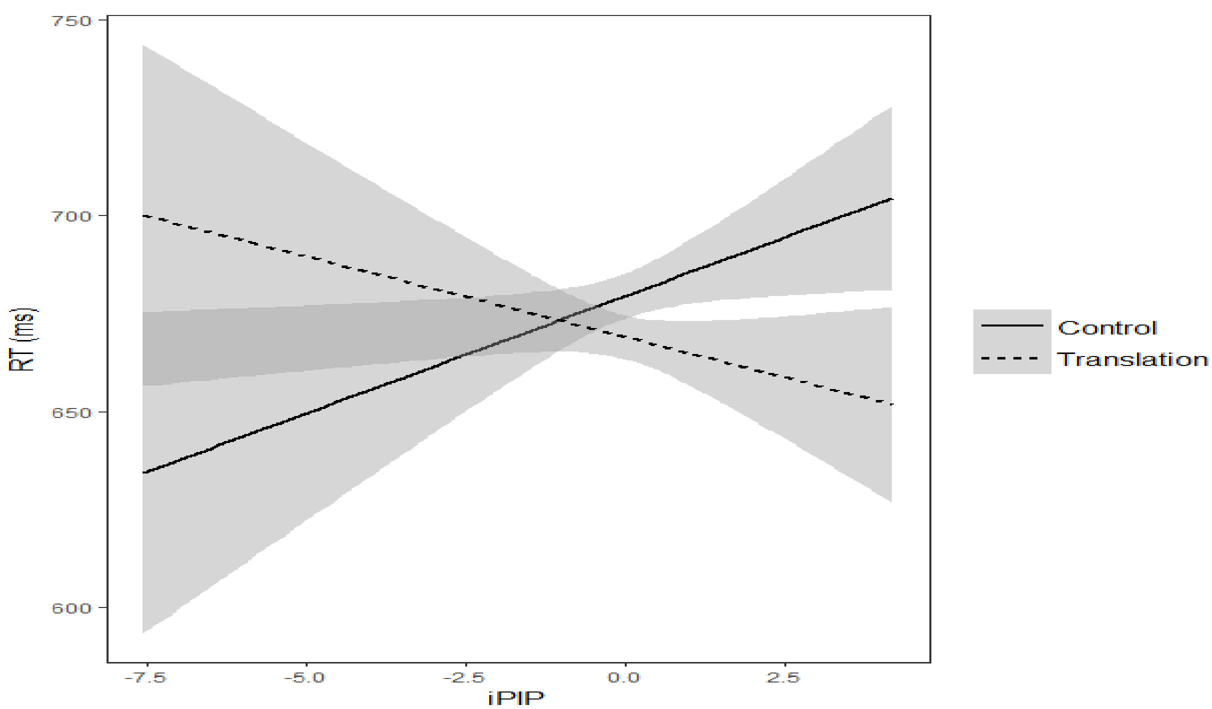

Figure 29. Response times as a function of prime and scaled iPIP, category typicality excluded, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Shaded areas represent 95% confidence intervals.

[27] The effect of iPIP was significant when the screening criteria were loosened, $t(10798) = 2.26$, $p = .024$.

5.1.3.2 Prime x iPIP Analysis, Typicality Included

The additive model was favoured over the interactive model, *BF* = 32.51, θ = 0.96. The effect of prime was significant in this analysis, β = -10.66, *SE* = 3.20, *t*(8550) = -3.35, *p* = .0009. Exemplar targets that were preceded by a translation prime (*M* = 662 ms) produced faster latencies than targets that were preceded by a control prime (*M* = 674 ms). Neither the effect of iPIP, *t* < 1, nor the two-way interaction were significant, β = -6.73, *SE* = 4.38, *t*(8550) = -1.54, *p* = .12[28]. The effects of prime and iPIP on RTs are shown in Figure 30. As shown in Figure 30, the joint effects of prime and iPIP trended towards an interaction, with larger priming effects being produced by high-frequency items that were more typical of the target category. This trend did not reach significance in the data, however.

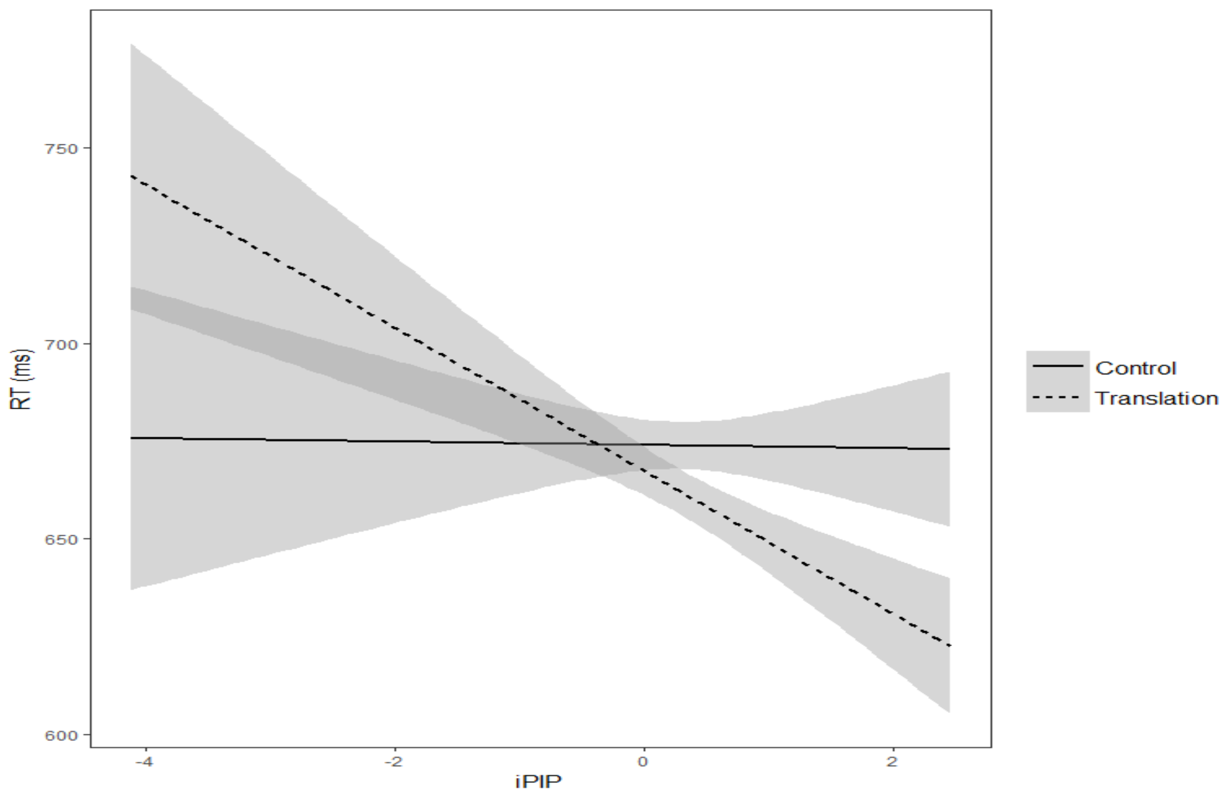

Figure 30. Response times as a function of prime and iPIP, category typicality included, combined Experiment 2 and 3 data, combined Experiment 2 and 3 coefficients. Shaded areas represent 95% confidence intervals.

[28] The two-way interaction between prime and iPIP was significant when the screening criteria were loosened, *t*(9565) = -2.04, *p* = .041.

5.1.3.3 Prime x PIP Analysis, Typicality Excluded

The interactive model was favoured over the additive model, *BF* = 10.72, θ = 383.24. This model involved a significant main effect of prime, β = -9.81, *SE* = 3.09, *t*(9444) = -3.17, *p* = .0015, and a two-way interaction between prime and PIP, β = -13.19, *SE* = 3.36, *t*(9444) = -3.92, *p* < .0001, which is shown in Figure 31. As shown in Figure 31, the combination of subject- and item-specific factors that were used to compute the sPIP and iPIP scores, as derived from the combined Experiment 2 and Experiment 3 data, significantly predicted priming effects in this combined data. Lower PIP scores were associated with an inhibitory effect of prime, while higher PIP scores were associated with a facilitative effect of prime.

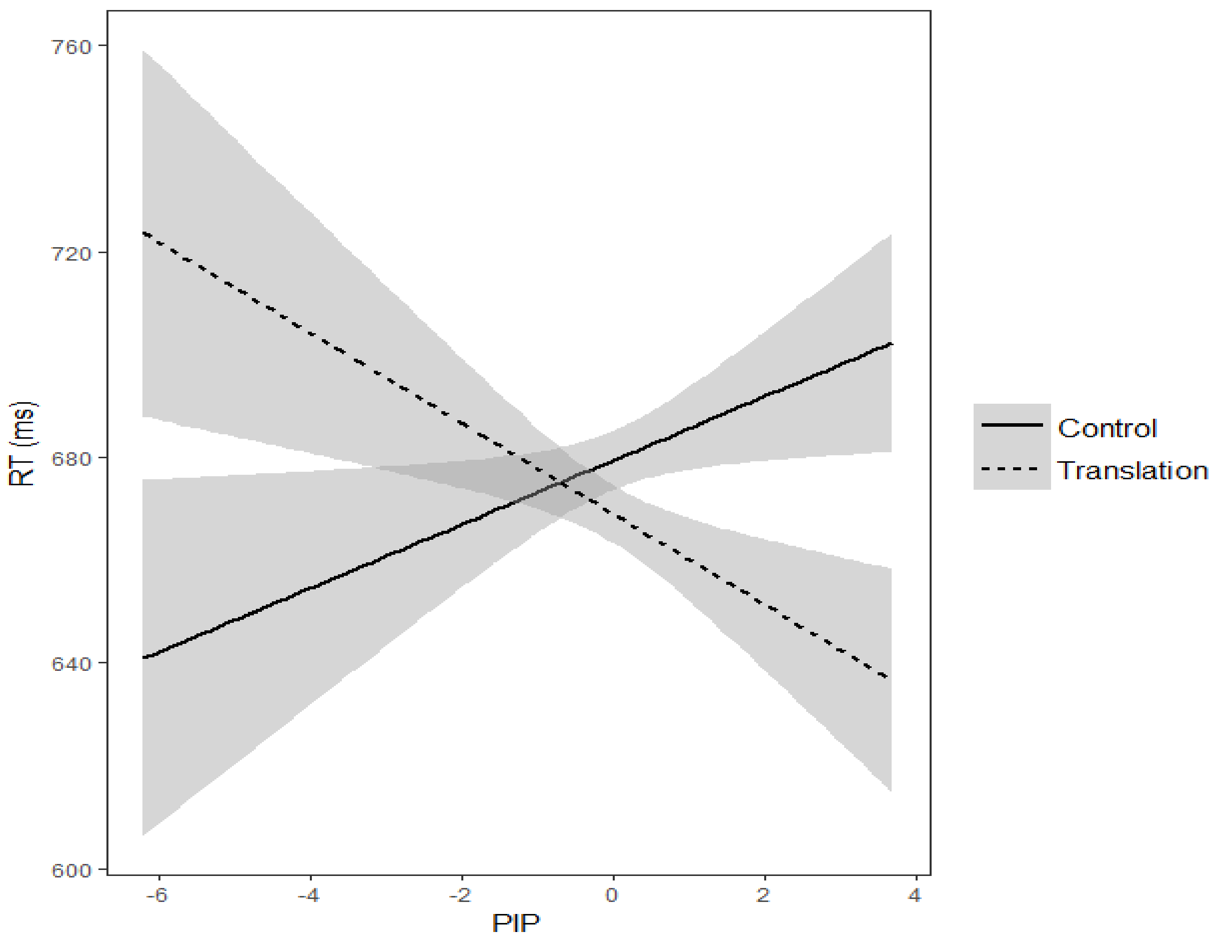

Figure 31. Response times as a function of prime and PIP, category typicality excluded, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Shaded areas represent 95% confidence intervals.

5.1.3.4 Prime x PIP Analysis, Typicality Included

The additive model was favoured by the Bayes factor, $BF = 1.67$, but not the relative likelihood, $\theta = 0.05$, meaning that the additive model was 1.67 times more likely to account for the data, but the interactive model was 20 times more likely to minimize the loss of information. The interactive model was thus favoured over the additive model, and involved a significant effect of prime, $\beta = -8.74$, $SE = 3.17$, $t(8550) = -2.76$, $p = .0058$. While the effect of PIP was nonsignificant, $\beta = -11.45$, $SE = 7.64$, $t(8550) = -1.50$, $p = .13$[29], the two-way interaction between prime and PIP was significant, $\beta = -10.76$, $SE = 3.51$, $t(8550) = -3.06$, $p = .0022$. This interaction is shown in Figure 32. Lower PIP scores were associated with an inhibitory effect of prime, while higher PIP scores were associated with a facilitative effect of prime.

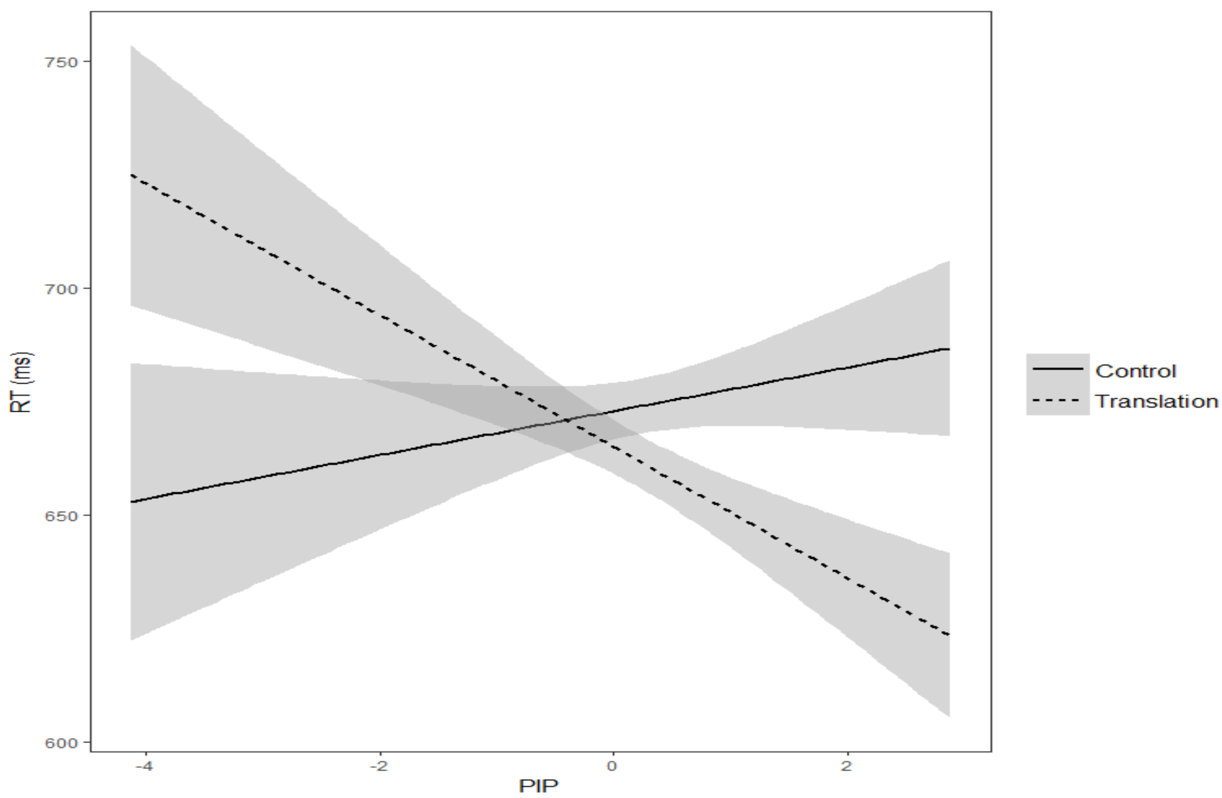

Figure 32. Response times as a function of prime and PIP, category typicality included, combined Experiment 2 and 3 data, combined Experiment 2 and 3 coefficients. Shaded areas represent 95% confidence intervals.

[29] The effect of PIP reached significance when the screening criteria were loosened, $t(9565) = -2.37$, $p = .018$.

5.1.3.5 Prime x List Analysis

The effect of prime was significant in this analysis, $\beta = -25.23$, $SE = 8.71$, $t(9444) = -2.90$, $p = .0038$, while the effect of list approached significance, $\beta = 25.72$, $SE = 13.48$, $t = 1.91$, $p = .056$. Response times were faster in List 1 ($M = 669$ ms) than they were in List 2 ($M = 678$ ms). Most importantly, the two-way interaction between prime and list was nonsignificant, $t < 1$. The priming effect was no larger in List 1 (11 ms) than it was in list 2 (9 ms).

5.1.4 Error Analysis, Experiment 2 Coefficients

5.1.4.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, $BF = 87.68$, $\theta = 2.36$. None of the effects were significant in any analysis, *z*s < 1.47, *p*s > .14. Mean error rates as a function of prime and sPIP tertile are shown in Figure 33.

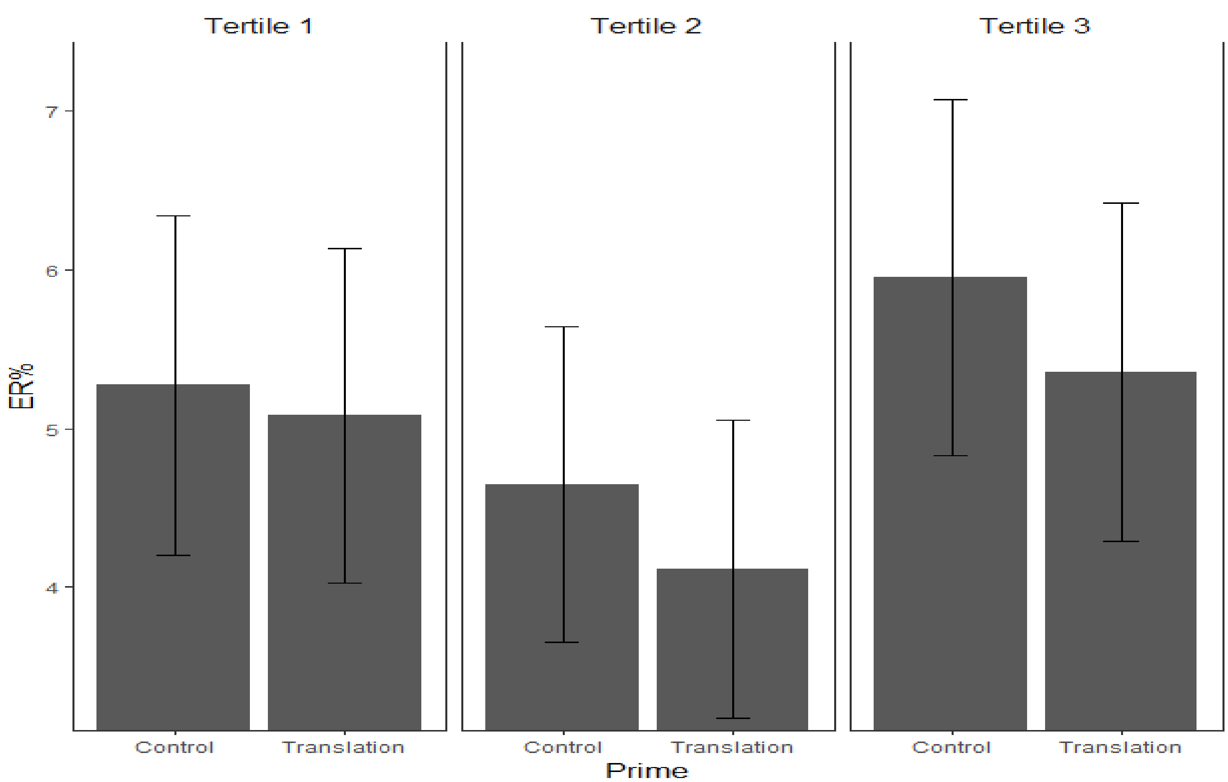

Figure 33. Mean error rates as a function of prime and sPIP tertile, combined Experiment 2 and 3 exemplar data, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

5.1.4.2 Prime x iPIP Analysis

The additive model was favoured over the interactive model, *BF* = 88.64, θ = 2.38, but again, none of the effects were significant in any analysis, *z*s < 1.20, *p*s > .23. Mean error rates as a function of prime and iPIP tertile are shown in Figure 34.

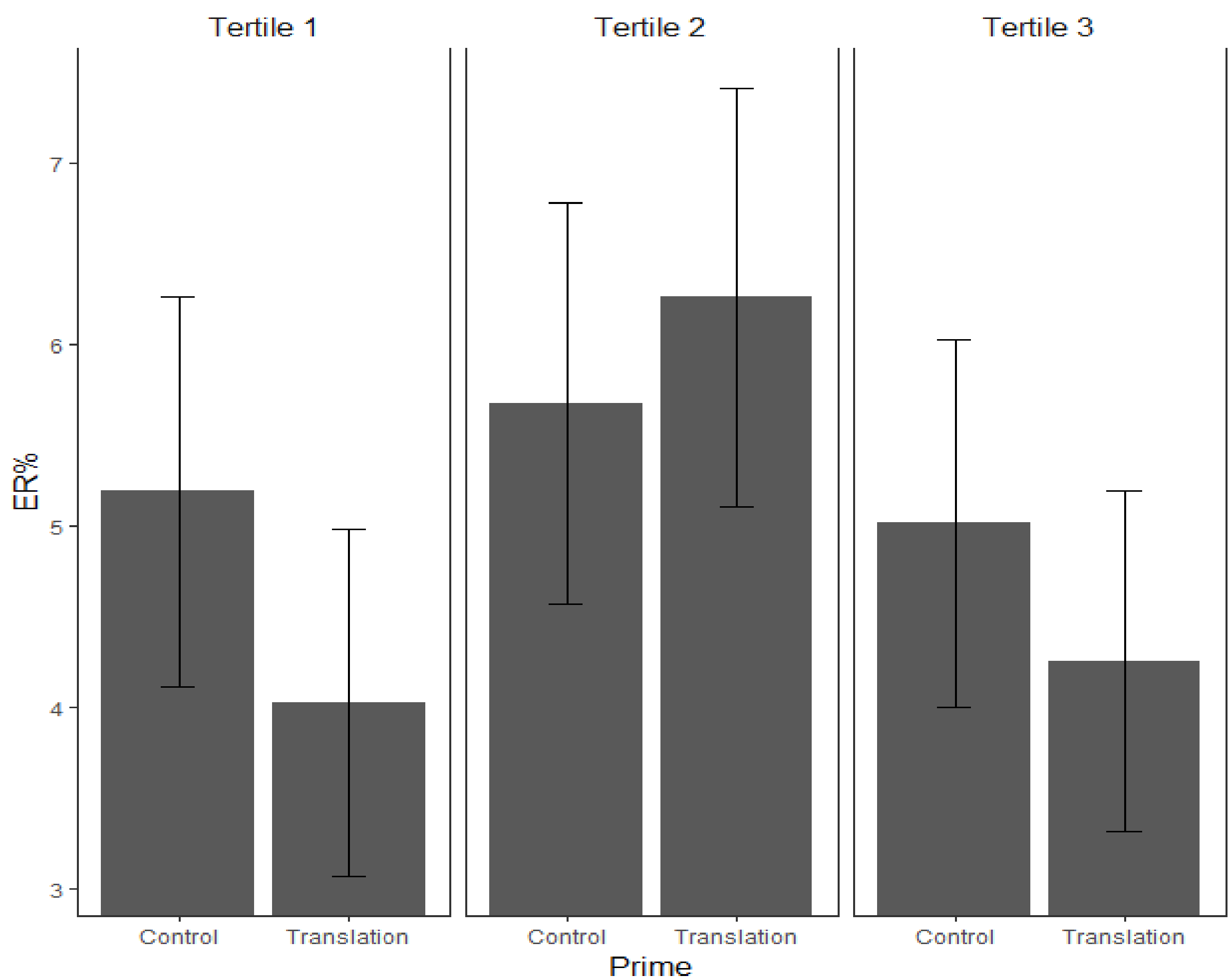

Figure 34. Mean error rates as a function of prime and iPIP tertile, combined Experiment 2 and 3 exemplar data, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

5.1.4.3 Prime x PIP Analysis

The additive model was favoured over the interactive model, *BF* = 75.77, θ = 2.04, and none of the effects were significant in any analysis, *z*s < 1. Mean error rates as a function of prime and PIP tertile are shown in Figure 35.

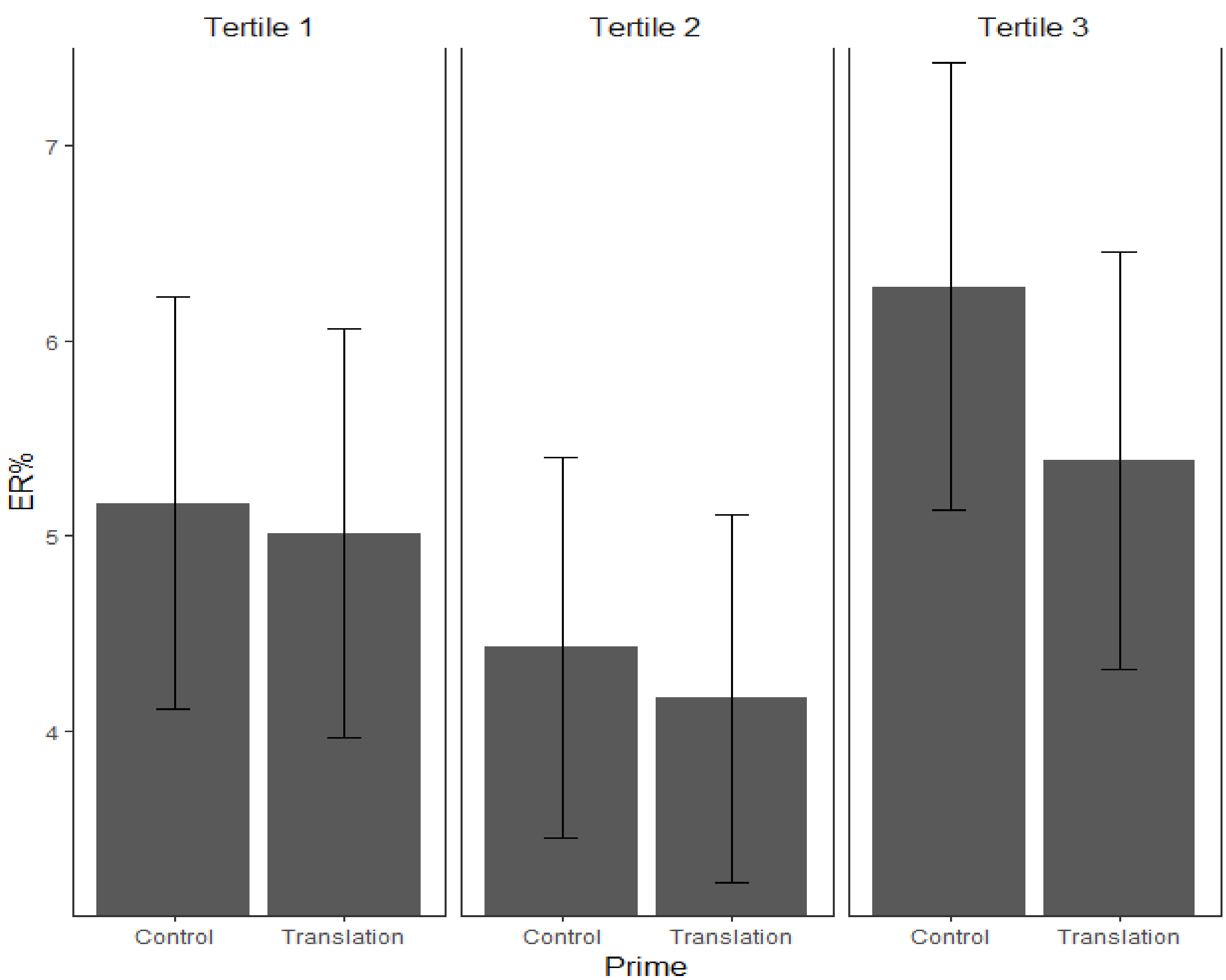

Figure 35. Mean error rates as a function of prime and PIP tertile, combined Experiment 2 and 3 exemplar data, Experiment 2 coefficients. Error bars represent 95% confidence intervals.

5.1.5 Error Analysis, Combined Coefficients

5.1.5.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, *BF* = 69.02, θ = 1.85, but none of the effects were significant in any analysis, *z*s < 1. Mean error rates as a function of prime and sPIP tertile are shown in Figure 36.

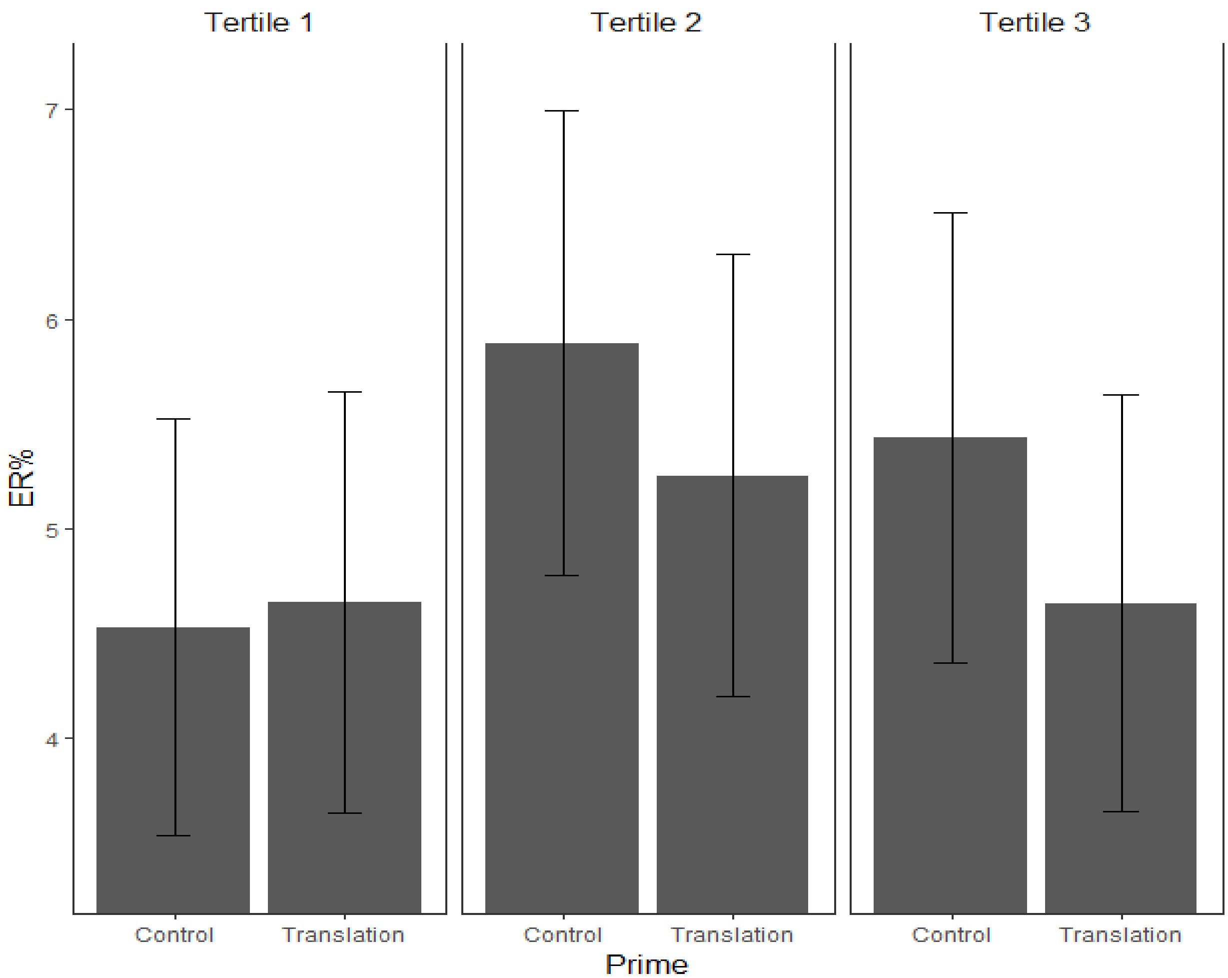

Figure 36. Mean error rates as a function of prime and sPIP tertile, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Error bars represent 95% confidence intervals.

5.1.5.2 Prime x iPIP Analysis, Typicality Excluded

The additive model was favoured over the interactive model, *BF* = 84.58, θ = 2.27, but again, none of the effects were significant in any analysis, *z*s < 1. Mean error rates as a function of prime and iPIP tertile are shown in Figure 37.

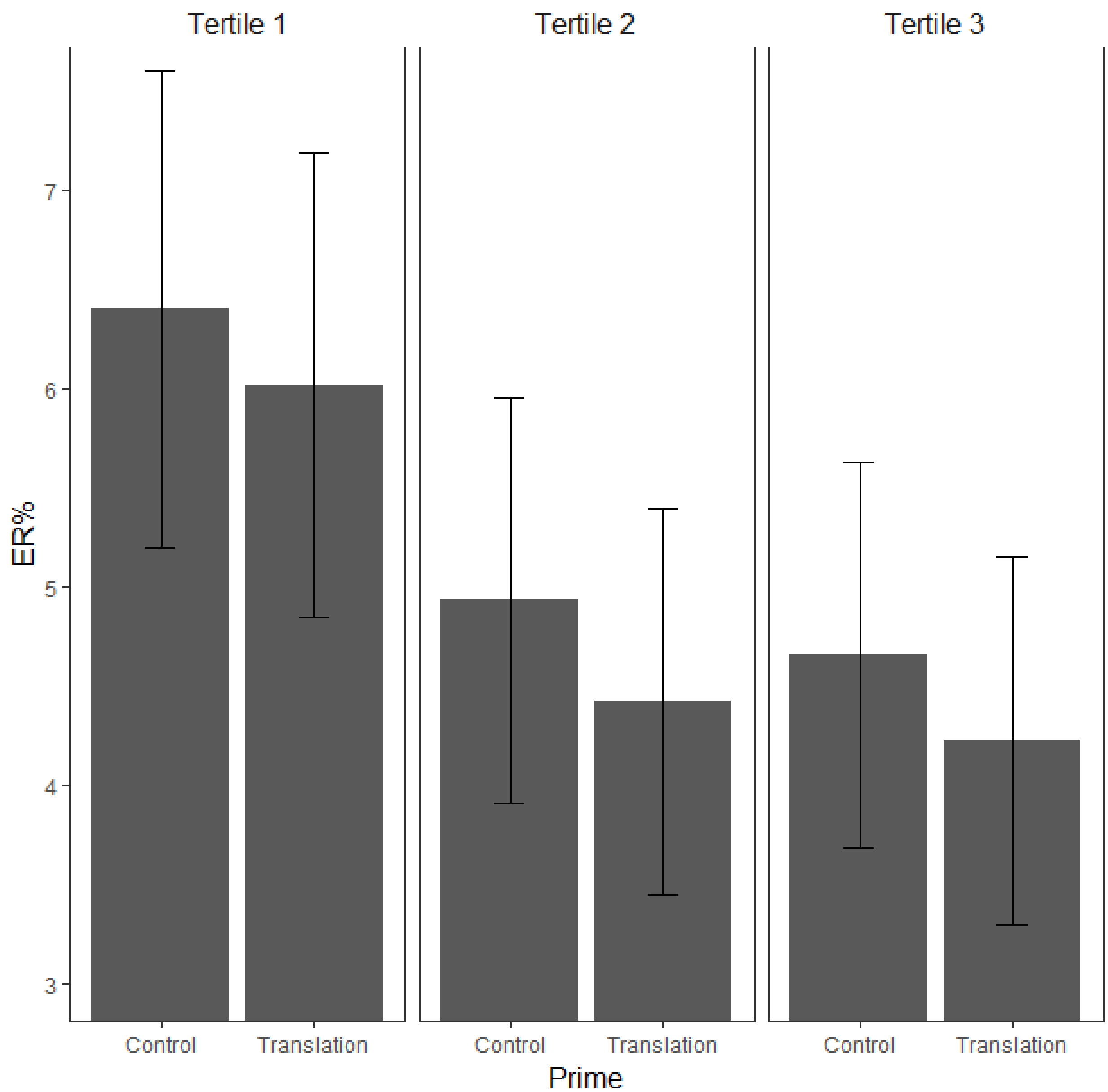

Figure 37. Mean error rates as a function of prime and iPIP tertile, category typicality excluded, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Error bars represent 95% confidence intervals.

5.1.5.3 Prime x iPIP Analysis, Typicality Included

The additive model was favoured over the interactive model, *BF* = 63.99, θ = 1.80. The only effect that approached significance was the effect of iPIP, $\beta = 0.16$, *SE* = 0.096, $z(9296) = 1.70$, $p = .09$. All other effects were nonsignificant, $z < 1$. Mean error rates as a function of prime and iPIP tertile for this analysis are shown in Figure 38.

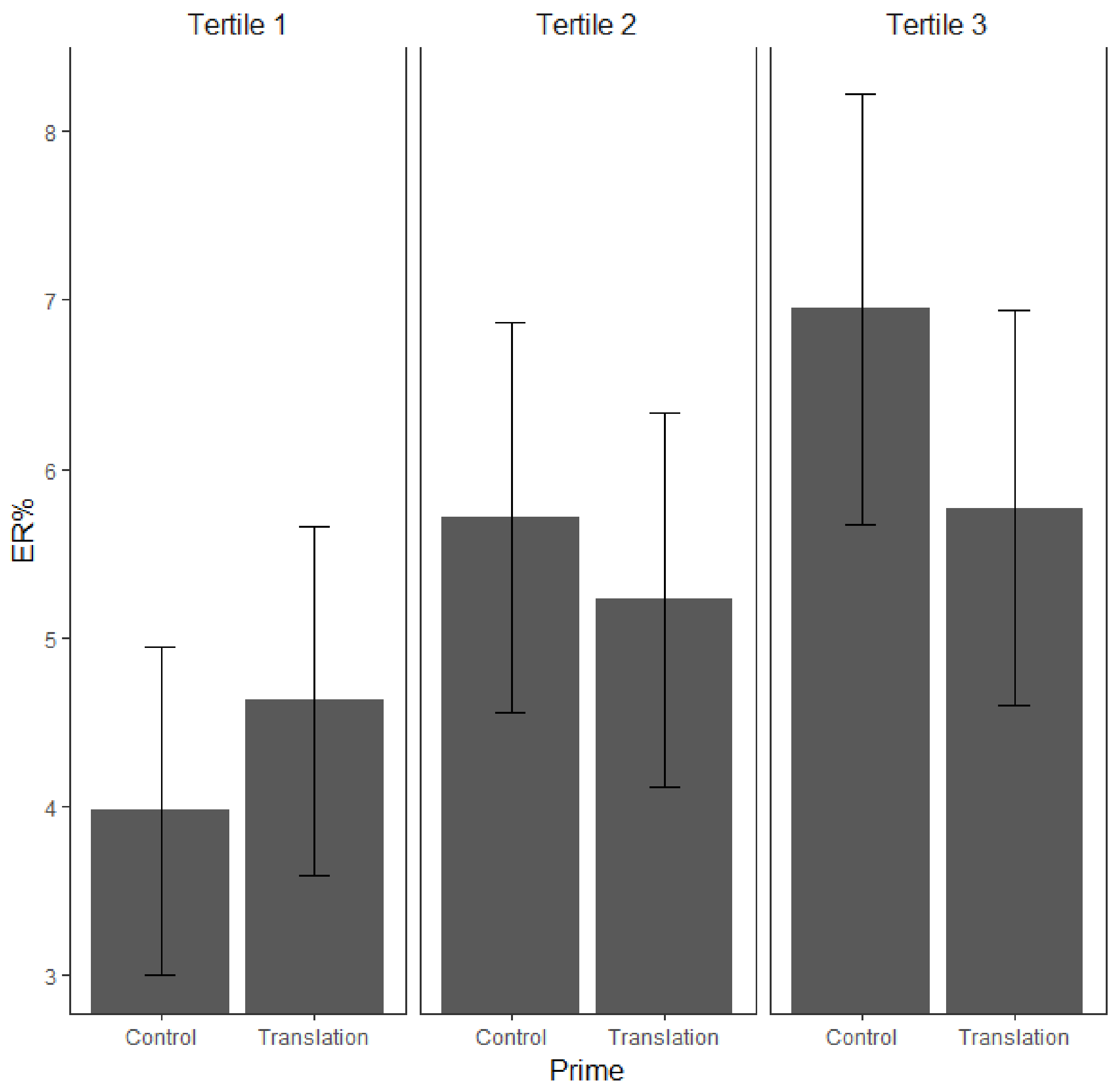

Figure 38. Mean error rates as a function of prime and iPIP tertile, category typicality included, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Error bars represent 95% confidence intervals.

5.1.5.4 Prime x PIP Analysis, Typicality Excluded

The additive model was again favoured over the interactive model, *BF* = 59.47, θ = 1.60, but again, none of the effects were significant in any analysis, *t*s < 1.08, *p*s > .28. Mean error rates as a function of prime and PIP tertile are shown in Figure 39.

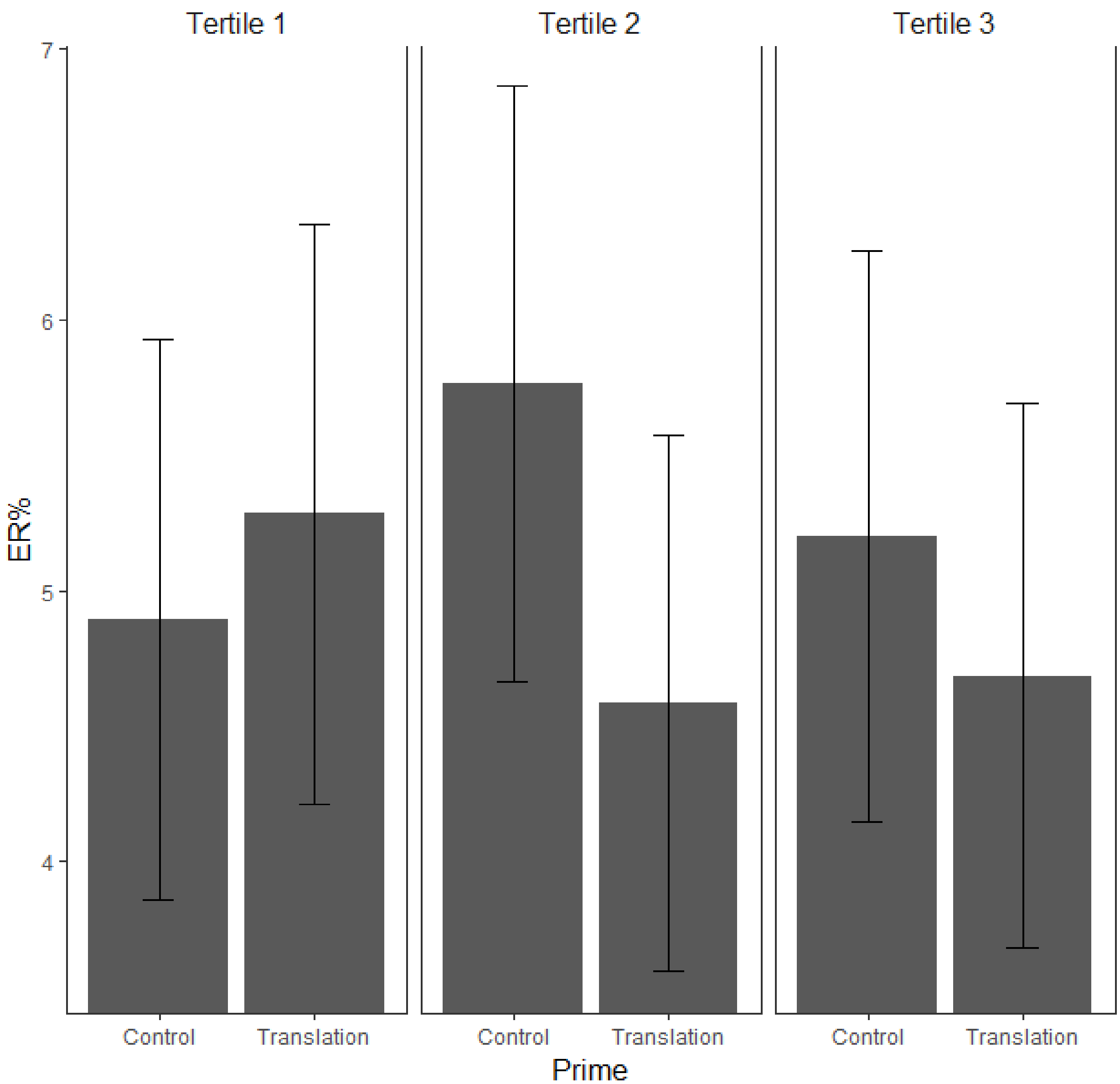

Figure 39. Mean error rates as a function of prime and PIP tertile, category typicality excluded, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Error bars represent 95% confidence intervals.

5.1.5.5 Prime x PIP Analysis, Typicality Included

The additive model was favoured over the interactive model, *BF* = 64.96, θ = 1.63, but again, none of the effects were significant, *z*s < 1.09, *p*s > .27. Mean error rates as a function of prime and PIP tertile are shown in Figure 40.

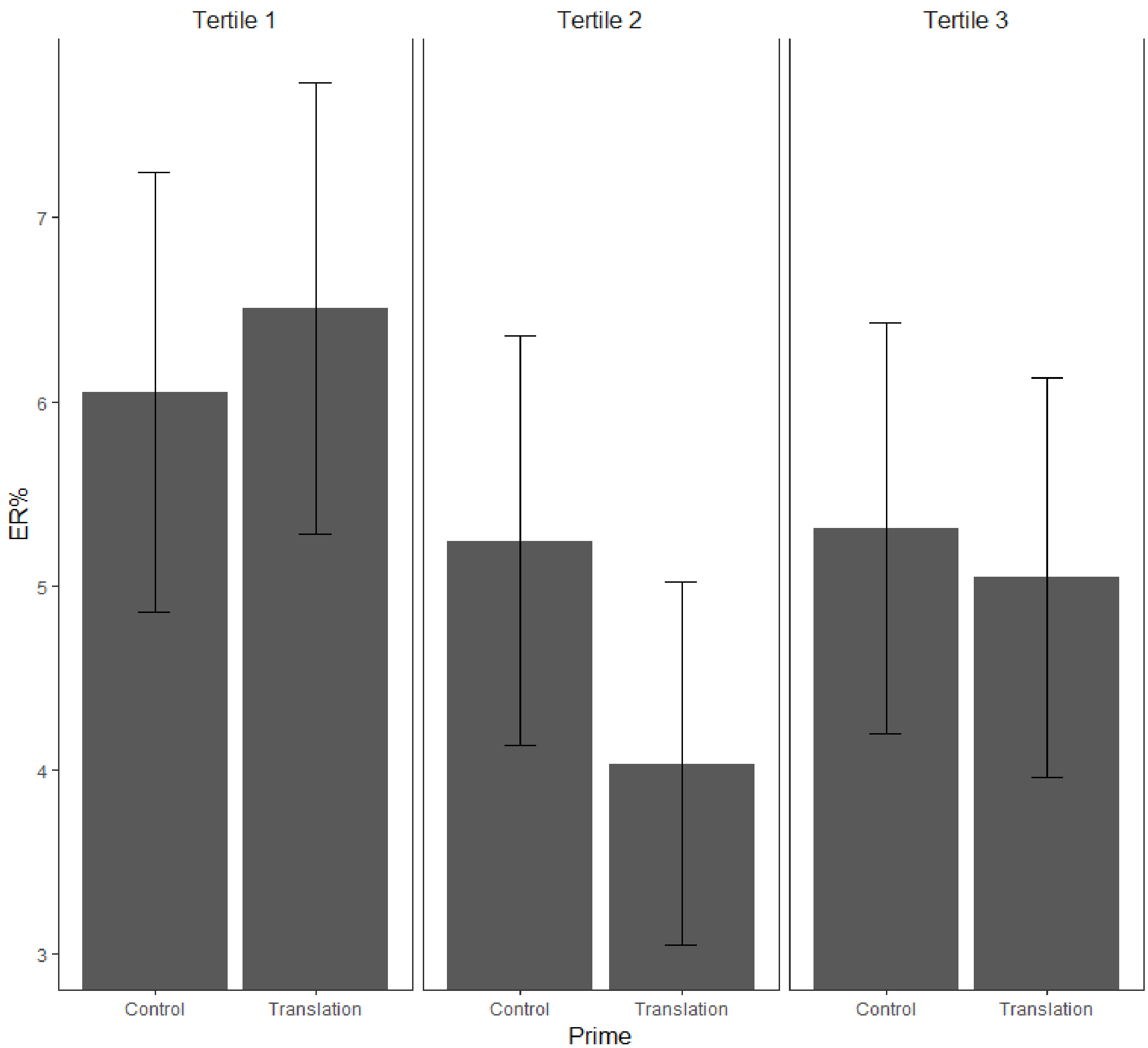

Figure 40. Mean error rates as a function of prime and PIP tertile, category typicality included, combined Experiment 2 and 3 exemplar data, combined Experiment 2 and 3 coefficients. Error bars represent 95% confidence intervals.

5.2 Discussion

Experiment 3 was conducted to test whether the results of Experiment 2 would be successfully replicated on a sample of subjects that were not used to construct the Experiment 2 PIP scores, that is, to test whether PIP predictions based on the Experiment 2 sample would generalize to other subjects. The findings of Experiment 3 have several implications. First, as with Experiment 2, Experiment 3 successfully replicated the significant effect of prime on RTs in semantic categorization that has been reported in prior research (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015). Second, Experiment 3 successfully demonstrated that the sPIP, iPIP, and PIP scores derived from Experiment 2 subjects could be used to make reasonable predictions for a new sample. Even when the coefficients derived from Experiment 2 were used, Experiment 3 still produced interactions between prime, sPIP, iPIP, and PIP in the exemplar RT data, indicating that a number of factors implicated in the sPIP and iPIP scores derived from Experiment 2 also predicted priming effects in new subjects. Deriving a new set of coefficients specifically from Experiment 3 data revealed several predictors that consistently predicted stronger priming effects in both Experiments 2 and 3. For sPIP, Experiment 3 implicated the percentage of English use in other social settings and at school as factors that predicted stronger priming effects, as well as English speaking proficiency, and Chinese listening proficiency, which directly replicated the sPIP coefficients derived from Experiment 2. Negatively associated with priming effects was Chinese reading proficiency, which was again replicated in Experiment 3. For iPIP, both Experiment 2 and Experiment 3 implicated the target's Google frequency (Tse et al., 2017), as well as the length of the prime.

There were a few differences in the variable coefficients derived from Experiments 2 and 3, however. In Experiment 2, English reading proficiency was not a significant predictive factor. In Experiment 3, this factor was a significant positive predictor. Further, in Experiment 2, the effects of English listening proficiency were positive, while the effects of Chinese writing proficiency were negative. The coefficients in Experiment 3 were in a different direction, as listening proficiency in English was a negative predictor, and writing proficiency in Chinese was a positive predictor. Overall, these findings suggest that there are individual differences in how these factors influenced processing in semantic categorization, and they were less reliable predictors overall than the use of English at school and other social contexts. In the item-based data, while Experiment 2 found a weak impact of prime CELEX frequency (higher frequency primes producing larger priming effects), this finding was not replicated by Experiment 3's results, which found that high-frequency L2 primes produced smaller priming effects than low-frequency L2 primes.

The overall analysis of the combined data confirmed that the most important facilitative subject-based factors in predicting priming effects were the percentage of time subjects used English at school and in other social contexts, while Chinese reading proficiency was the most reliable negative subject-based predictor of priming effects. The combined analysis also confirmed that target frequency was the most important positive item-based factor in predicting priming effects. Priming effects were larger for high-frequency targets than they were for low-frequency targets in Experiments 2 and 3. Prime CELEX frequency and target stroke count were negative predictors, in that priming effects were smaller for targets preceded by high-frequency English translation primes, and when the targets had a large number of strokes, replicating the results of Experiment 3. Once again, the combined analysis of Experiments 2 and 3 demonstrate that the effect of factors such as target frequency on translation priming is task-dependent. In lexical decision, priming effects were smaller for high-frequency targets than low-frequency targets, while in semantic categorization, priming effects were larger for high-frequency targets than low-frequency targets.

Finally, the combined analysis of Experiments 2 and 3 found that L2-L1 translation priming was affected by the typicality of the English exemplar prime. Priming effects tended to be larger when the translation prime was a more typical representation of the category than when the prime was an atypical member of the category. This effect did not reach significance when the initial screening criteria were set, but still trended towards an interaction. When the screening criteria were loosened, however, this interaction reached significance.

Overall, the results of Experiments 2 and 3 are consistent with the notion that priming effects in the semantic categorization task are largely predicted by the extent to which bilinguals actively use their L2 in the social environments that they encounter on a daily basis, and the effect is larger when the exemplar targets are high frequency, and their English translation equivalents are highly typical members of the category, perhaps suggesting that the targets need to be more frequently encountered and more typical members of the category. A full discussion of the interpretations and implications of these results can be found in the General Discussion.

Chapter 6

6 Experiment 4

An additional purpose of the present research was to address the discrepancy between the assumptions of the Episodic L2 Hypothesis (Jiang & Forster, 2001), and the empirical results from prior studies that have shown significant L2-L1 translation priming effects in tasks that are assumed to tap into lexical and semantic (as opposed to episodic) memory, in particular, the semantic categorization task (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), and the lexical decision task for highly proficient bilinguals (e.g., Nakayama et al., 2016). As was discussed previously, if L2 primes cannot activate lexical or semantic representations of L1 targets because L2 words are represented in a different memory system than L1 words, then one would not expect to find priming effects in either task, and yet empirical results from both the present research and from prior research have produced evidence contrary to this prediction. However, as was also discussed, the Episodic L2 Hypothesis could be amended to account for these apparently contradictory results if its assumptions were changed slightly. First, consistent with the original account, it is assumed that L2 words are initially represented in episodic memory rather than lexical memory. However, over the course of acquiring greater knowledge about one's L2 and becoming more proficient in the language, the locus of representation qualitatively shifts from an episodic representation to a lexical representation, as processing in L2 becomes more efficient and automatized. This shift can be proposed to be affected by both learner- and word-level factors. Learner-level factors would include factors such as global L2 proficiency, as well as subfactors such as speaking, reading, writing, listening proficiency, vocabulary size, the age at which learners acquired their L2, and the amount of time that the learner has been learning their L2. Word-level factors would include factors such as word frequency and familiarity. Such an amendment could potentially account for at least some of the contradicting findings of prior research, while providing a plausible account of how the memory systems used in processing language change over the course of knowledge acquisition.

To examine these ideas, a speeded episodic recognition task was used. If L2 knowledge is initially represented in episodic memory, but shifts to lexical memory, potentially on a word-by-word basis, over the course of acquiring greater knowledge and skill in one's L2, then two predictions can be made. First, for subjects who are less proficient in their L2, a significant priming effect should arise in this task. However, for subjects that are highly proficient, one consequence of acquiring more L2 proficiency would be that the translation prime would no longer reliably facilitate the recognition of old items. In fact, having a high degree of L2 proficiency may make the task more difficult by increasing the feeling of familiarity for primed new items, and cause an inhibitory effect to arise.

Under the circumstance where no priming effect is obtained in Experiment 4, follow-up analyses were conducted to test whether the null priming effect was due to fatigue effects from doing a long, taxing experiment. Experiment 4 used a large number of stimuli to achieve statistical power, and the task was divided into multiple blocks. Because this task was longer than the task used by Forster and colleagues (Jiang & Forster, 2001; Witzel & Forster, 2012), the risk of fatigue effects was higher. Such fatigue effects should not occur in the first block of the experiment, however. As such, follow-up analyses were conducted on the first block of the experiment in circumstances where the priming effect was nonsignificant.

6.1 Method

6.1.1 Subjects

Subjects were 44 students (28 female, 16 male) at the University of Western Ontario. Thirty of these subjects completed the study for course credit, while the remaining 14 subjects were provided monetary compensation. Subjects ranged between 18 to 30 years of age (M = 21.13, SD = 3.34). Forty-three of these subjects were right-handed, and only one subject reported being left-handed. All subjects had normal or corrected-to-normal vision. Out of the 44 subjects that participated, 37 reported speaking Mandarin and English. In addition, seven subjects reported being trilingual, with one participant speaking Mandarin, English, and Japanese, one participant speaking Mandarin, English, and Spanish, and five participants speaking Cantonese, Mandarin, and English. Thus, six of the 44 participants in this experiment could read in additional orthographic systems, with five participants being able to read Traditional Chinese script, and one participant being able to read Japanese kana and Kanji.

6.1.2 Stimuli

A set of 480 words were used in Experiment 4. Some of these words were derived from Experiment 1's stimulus set. All words were composed of two characters, and targets were either primed by a translation prime, or by an unrelated prime. Experiment 4 was counterbalanced using eight lists. The purpose of using eight lists was to use a large sample of stimuli for testing. However, having 480 stimuli on a single list was very time-consuming, so the stimuli that participants were presented varied by list. Half of the words appeared on Lists 1-4, while the other 240 words appeared on Lists 5-8. On each list, half of the words appeared during the initial study phase, and half of the targets appeared as new targets. In addition, half of the targets in both the Old and New conditions were preceded by a translation prime, and half were preceded by a control prime. Each word appeared both as an old and a new target, and with both a control and translation prime across all lists. The mean Google frequency and stroke count of the targets can be found in Table 2. All words used in Lists 1-4 of Experiment 4 can be found in Appendix D, while all words used in Lists 5-8 of Experiment 4 can be found in Appendix E.

6.1.3 Measures

The same measures that were used in Experiments 1-3 were included in Experiment 4, with a few additions. First, subjects were also assessed on what age they first acquired English. Second, based on this information, the approximate amount of time that subjects had been learning English was estimated. Both of these factors were included in the computation of sPIP, iPIP, and PIP for Experiment 4.

6.1.4 Procedure

The procedure was a modified version of Jiang and Forster's (2001) speeded episodic recognition task, using three training-testing phases as opposed to one. This task involved two phases. First, in a study phase, subjects were presented 40 Chinese words to study and memorize. At first, each word was presented individually on a computer screen for 2 seconds, with a 1 second interval between presentations. The 40 words were cycled through twice in this manner, so subjects saw each word twice. Afterwards, the words were then presented in five sets of eight words. Subjects were given the opportunity to take as long as they wanted to memorize the words in each set, and then they could press the spacebar to advance to the next set. After completing every set, all of the words were presented together once more for a final review. Subjects could review these words for as long as they wanted to before advancing to the testing phase. Subjects were then told that a memory test would be given, and they were asked to remember as many of the words that they were presented as possible.

During the testing phase, subjects were instructed to decide as quickly but as accurately as possible whether the word presented on the screen was one of the words that they had studied during the training phase by either pressing the *?* key if the target was a word that was presented during the training phase, or the *z* key if the word was not studied previously. Each testing phase consisted of 80 words, half of which were presented during the training phase, and half of which were new. Upon the completion of a testing phase, subjects were given the opportunity to take a break. Once they were ready, they began another training-testing cycle, which included a new set of 40 words for them to memorize. In total, subjects completed three training-testing phases.

6.2 Results

6.2.1 Data Trimming

The data were trimmed using the same method as in Experiments 1-3. In the first phase of the trimming, one item (0.20% of the total data) and two subjects (4.54% of the total data) were removed. In the second phase, 11 items (2.05% of the total data), and three subjects (6.80% of the total data) were removed. Finally, errors (9.62% of the total data), and response times that exceeded 3.5 standard deviations from each subject’s mean, or were faster than 250 ms and slower than 2000 ms were removed (1.57% of the total data). In total, 24.77% of the data was removed in Experiment 4[30].

[30] In follow-up analyses with loosened criteria, the Mahalanobis distance criterion was loosened to .001 and outliers were screened if they deviated from each subject’s mean by 3 standard deviations, or were faster than 200 ms and slower than 3000 ms (2.33% of the data). Doing so resulted in no subjects or items being screened as multivariate outliers. All other data loss was due to participants and items being excluded for having error rates exceeding 50% (4.73% of the data) and from the exclusion of errors (12.09% of the data). Eighty-one percent of the data was retained in this analysis using these screening criteria.

6.2.2 PIP

The PIP coefficients for Experiment 4 can be found in Table 12. For Experiment 4, positive sPIP coefficients included English writing, speaking reading, and listening proficiency, Chinese writing and speaking proficiency, the percentage of English use in other social contexts, and the number of years that the subject has been learning English. Negative sPIP coefficients included Chinese reading and listening proficiency, and the age at which the subject first learned English. For iPIP, there were no positive coefficients. The predictor with the largest negative effect on priming effects was the number of strokes that the target was composed of, followed by the target's frequency, and the prime's CELEX frequency and length.

Table 12. *PIP Coefficients for Experiment 4.*

	PIP Coefficient Values
sPIP	
CR	-3.39
FL	-2.23
CL	-1.81
EL	1.32
CS	2.04
ER	2.07
YL	2.45
ES	2.46
PEO	2.57
CW	2.57
EW	2.86
iPIP	
L	-2.02
PCEL	-3.57
GF	-9.68
NS	-16.97

Note: CR = Self-reported Chinese reading proficiency; FL = Age at which subject first learned English; CL = Self-reported Chinese listening proficiency; EL = Self-reported English listening proficiency; CS = Self-reported Chinese speaking proficiency; ER = Self-reported English reading proficiency; YL = Number of years that subject has been learning English; ES = Self-reported English speaking proficiency; PEO = Percentage of time English is spoken in social settings outside of the home and school; CW = self-reported Chinese writing proficiency; EW = Self-reported English writing proficiency; L = Prime length; PCEL = Prime CELEX frequency; GF = Target Google frequency; NS = Number of strokes.

6.2.3 Reaction Time Analysis, Full Data

6.2.3.1 Old Trials Analysis

6.2.3.1.1 Prime x sPIP Analysis

While the additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 10.87, the interactive model was favoured over the additive model in the relative likelihood analysis, θ = 2.06. When compared to a restricted model which excluded the main effect of prime, and retained the effect of sPIP and the two-way interaction, however, the restricted model was favoured over the additive model in both analyses, *BF* = 5.52, θ = 5.52, indicating that the reason the additive model was favoured over the interactive model was because of the inclusion of prime as a main effect, not because of the inclusion of the two-way interaction. As such, the interactive model was selected over the additive model. This model found no main effect of prime, $t < 1$. Targets that were preceded by a translation prime (*M* = 696 ms) and targets that were preceded by a control prime (*M* = 690 ms) produced similar latencies. While the effect of sPIP was nonsignificant, *t*s < 1, there was a marginally significant two-way interaction between prime and sPIP, $\beta = 4.71$, $SE = 2.54$, $t(3709) = 1.85$, $p = .064$, which is shown in Figure 41. The effect of sPIP on RTs varied as a function of the prime which preceded the target. When the prime was a translation prime, higher sPIP scores were associated with faster RTs than lower sPIP scores. When the prime was a control prime, however, sPIP had no effect on RTs. The result was an interaction. Overall, priming effects were larger for subjects who reported higher global proficiency in English, reported using English more in other social contexts, reported learning English for a longer period of time and acquired English at a younger age, and who had higher writing and speaking proficiency, but relatively lower reading and listening proficiency in Chinese.

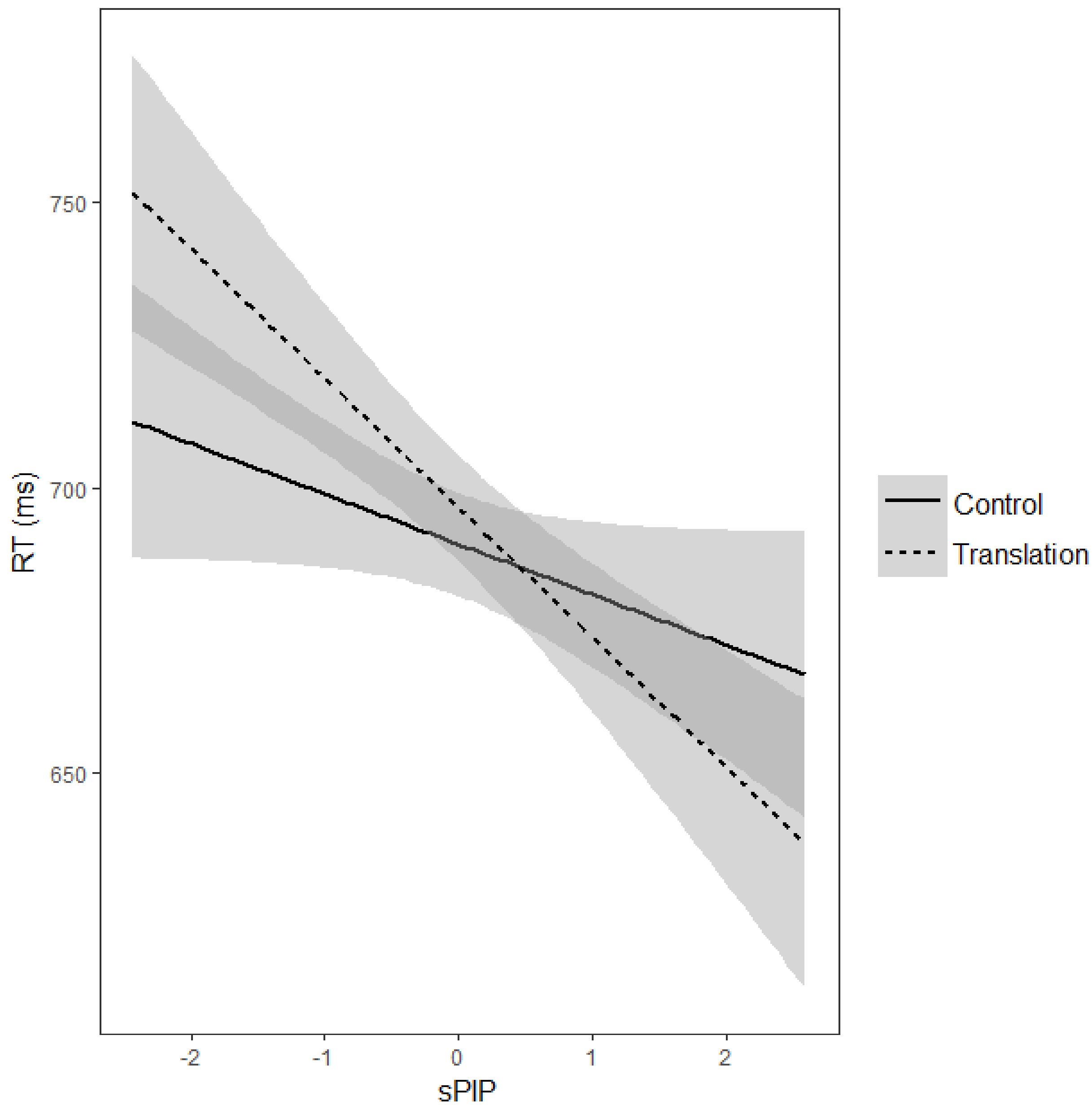

Figure 41. Response times as a function of prime and scaled sPIP, Experiment 4 Old trials. Shaded areas represent 95% confidence intervals.

6.2.3.1.2 Prime x iPIP Analysis

The additive model was favoured over the interactive model, *BF* = 20.06, θ = 0.90, and involved a significant effect of iPIP, $\beta = -14.30$, $SE = 5.10$, $t(3709) = -2.80$, $p = .0051$, but neither the effect of the prime, $t < 1$, nor the two-way interaction were significant, $\beta = 3.16$, $SE = 2.44$,

$t(3709) = 1.29$, $p = .20$[31]. As seen Figure 42, higher iPIP scores were associated with faster RTs overall, but priming had little impact on RTs overall. Numerically, priming effects were larger for low-frequency Chinese targets with relatively fewer strokes, which were preceded by shorter, lower-frequency English primes, but this trend was nonsignificant.

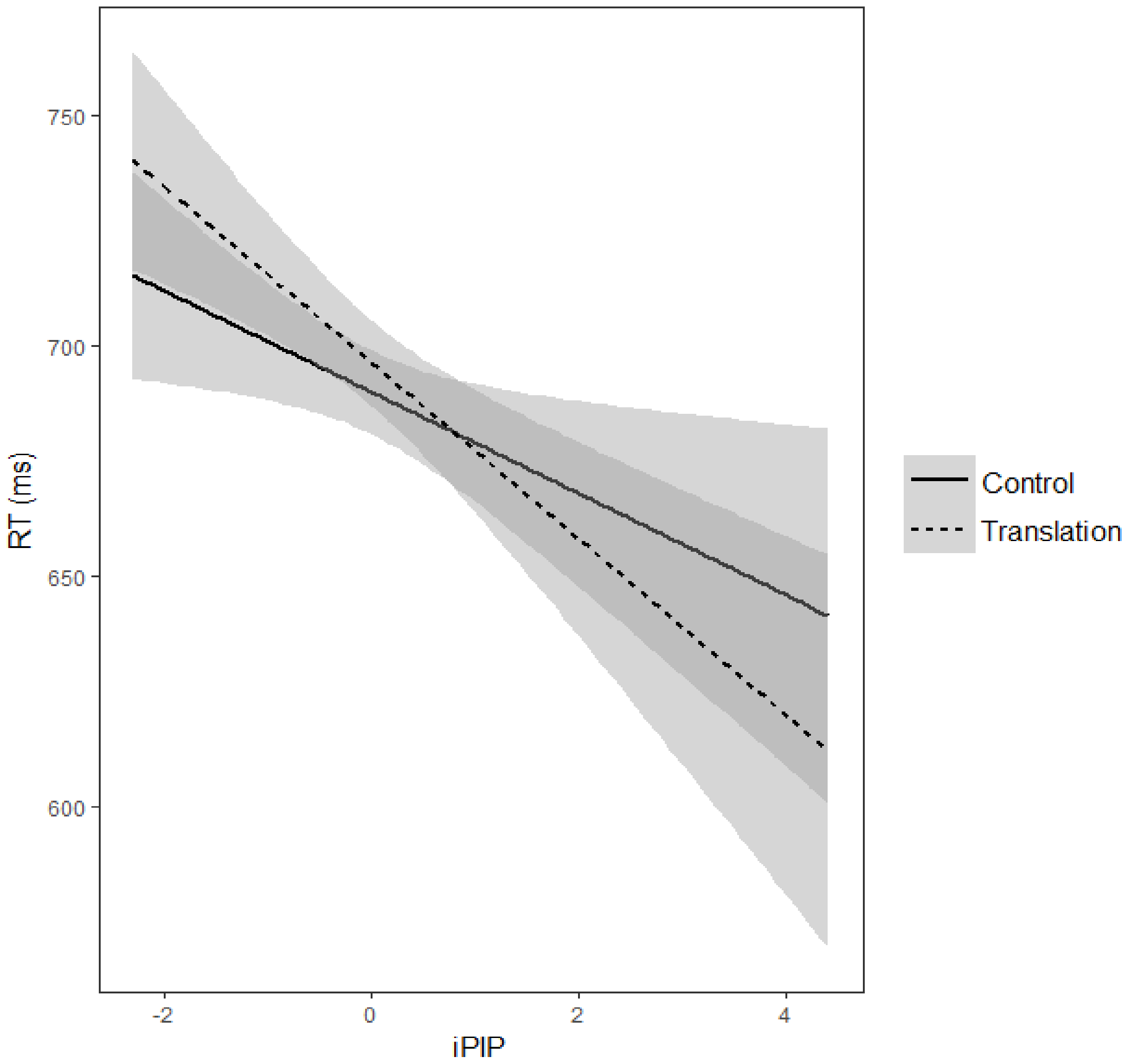

Figure 42. Response times as a function of prime and scaled iPIP, Experiment 4 Old trials. Shaded areas represent 95% confidence intervals.

[31] The effect of iPIP was nonsignificant when the screening criteria were loosened, $t(4015) = -1.31$, $p = .19$, and the two-way interaction between prime and iPIP was marginally significant when the screening criteria were loosened, $t(4015) = -178$, $p = .075$.

6.2.3.1.3 Prime x PIP Analysis

While the additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 10.54, the interactive model was favoured over the additive model in the relative likelihood analysis, $\theta = 2.13$. A follow-up comparison using a restricted model which excluded the effect of the prime, and retained the effect of PIP and the two-way interaction between prime and PIP showed that this model was favoured over the additive model in both analyses, *BF* = 5.71, θ = 5.71, indicating that the reason that the additive model was favoured over the interactive model was because the interactive model included the main effect of the prime, not because the interactive model included the interaction term. As such, the interactive model was selected over the additive model. While the main effect of the prime was nonsignificant in this analysis, $t < 1$, both the effect of PIP, $\beta = -16.92$, *SE* = 8.83, $t(3709) = -1.92$, $p = .055$, and the two-way interaction between prime and PIP approached significance, $\beta = 4.67$, *SE* = 2.48, $t(3709) = 1.88$, $p = .06$[32]. As shown in Figure 43, the effect of PIP on RTs varied as a function of the prime that the target was preceded by. When preceded by a translation prime, larger PIP scores were associated with faster RTs. When preceded by a control prime, the effects of PIP on RTs were relatively smaller. As a result, an inhibitory effect of the prime emerges at lower PIP scores, and a facilitative effect of the prime emerges at higher PIP scores. In sum, the combined subject- and item-specific factors that were included in the computation of the sPIP and iPIP scores predicted larger priming effects in Experiment 4.

[32] The effect of PIP was nonsignificant when the screening criteria were loosened, $t(4015) = -1.21$, $p = .23$, and the two-way interaction between prime and PIP was significant, $t(4015) = -2.17$, $p = .03$.

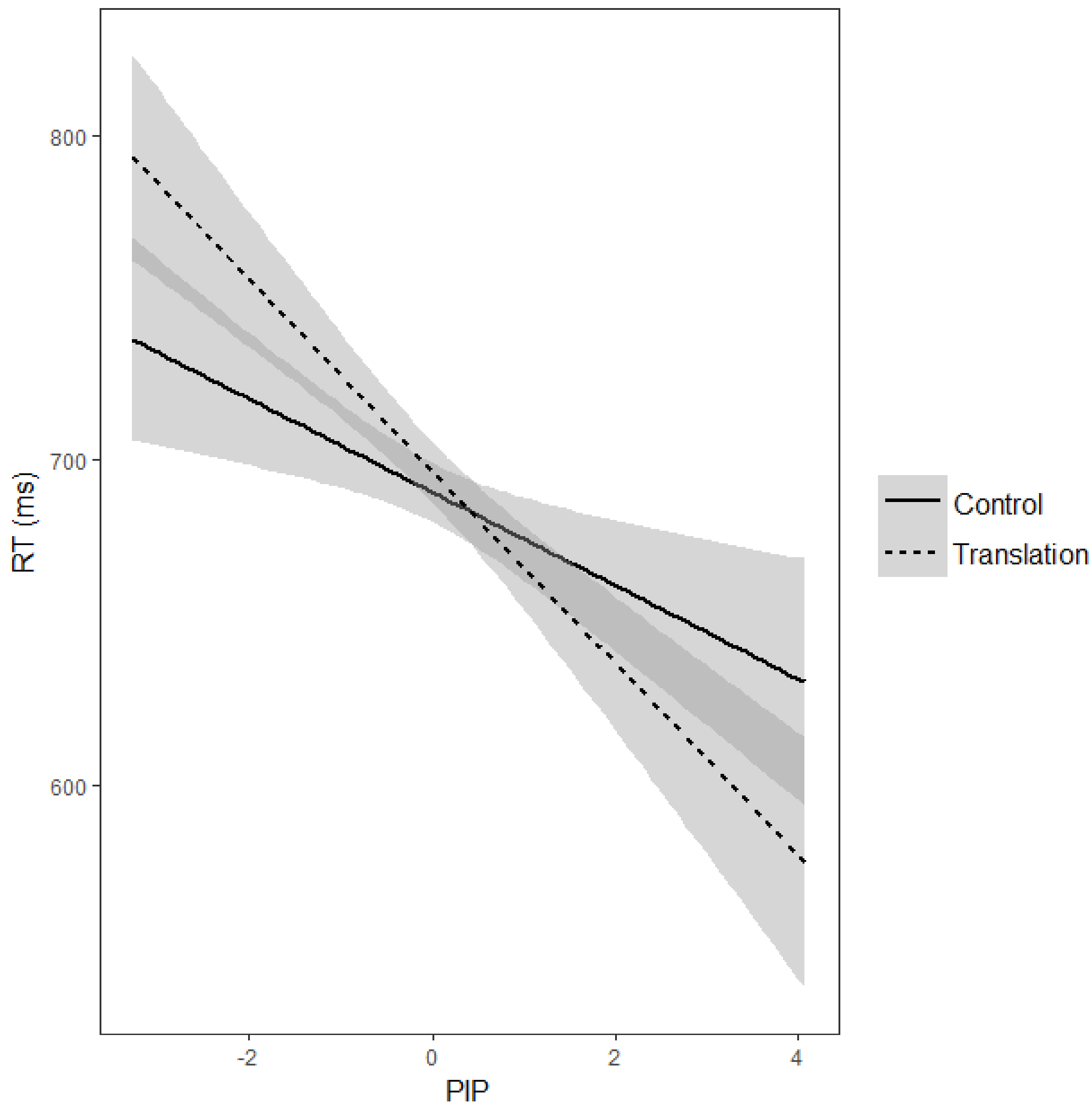

Figure 43. Response times as a function of prime and PIP, Experiment 4 Old trials. Shaded areas represent 95% confidence intervals.

6.2.3.1.4 Prime x Order Analysis

None of the effects were significant in this analysis, *t*s < 1.02, *p*s > .30[33].

[33] The prime x list analyses would not converge, likely because the number of items per cell across 8 lists was relatively small.

6.2.3.2 New Trials Analysis

6.2.3.2.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, *BF* = 62.93, θ = 2.68, but none of the effects were significant in any analysis, *t*s < 1. The effects of prime and sPIP on the RTs of New trials are shown in Figure 44.

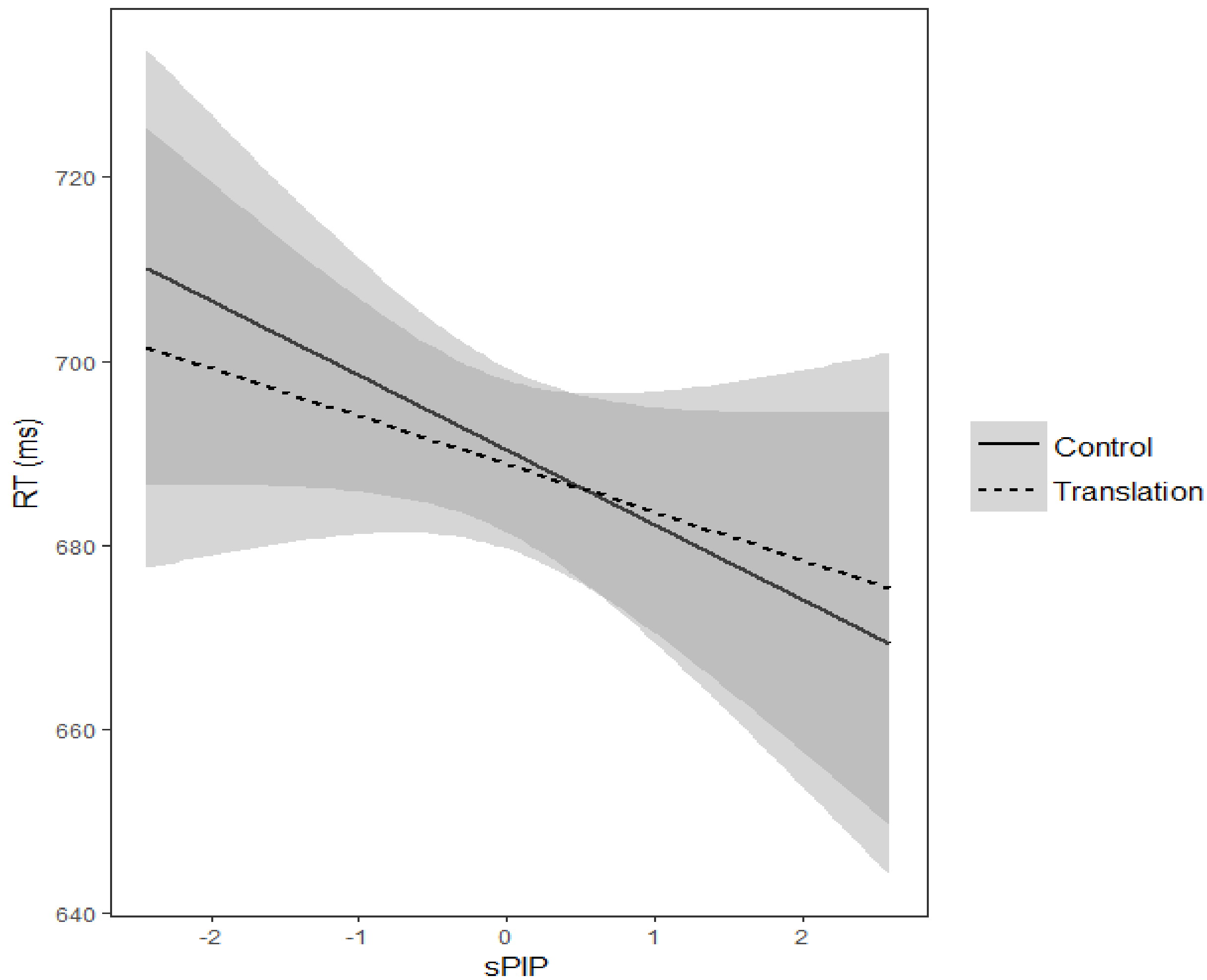

Figure 44. Response times as a function of prime and scaled sPIP, Experiment 4 New trials. Shaded areas represent 95% confidence intervals.

6.2.3.2.2 Prime x iPIP Analysis

The additive model was again favoured over the interactive model, *BF* = 44.24, θ = 1.88. Again, none of the effects were significant in any analysis, *t*s < 1.16, *p*s > .24. The effects of prime and iPIP on the RTs of New trials are shown in Figure 45.

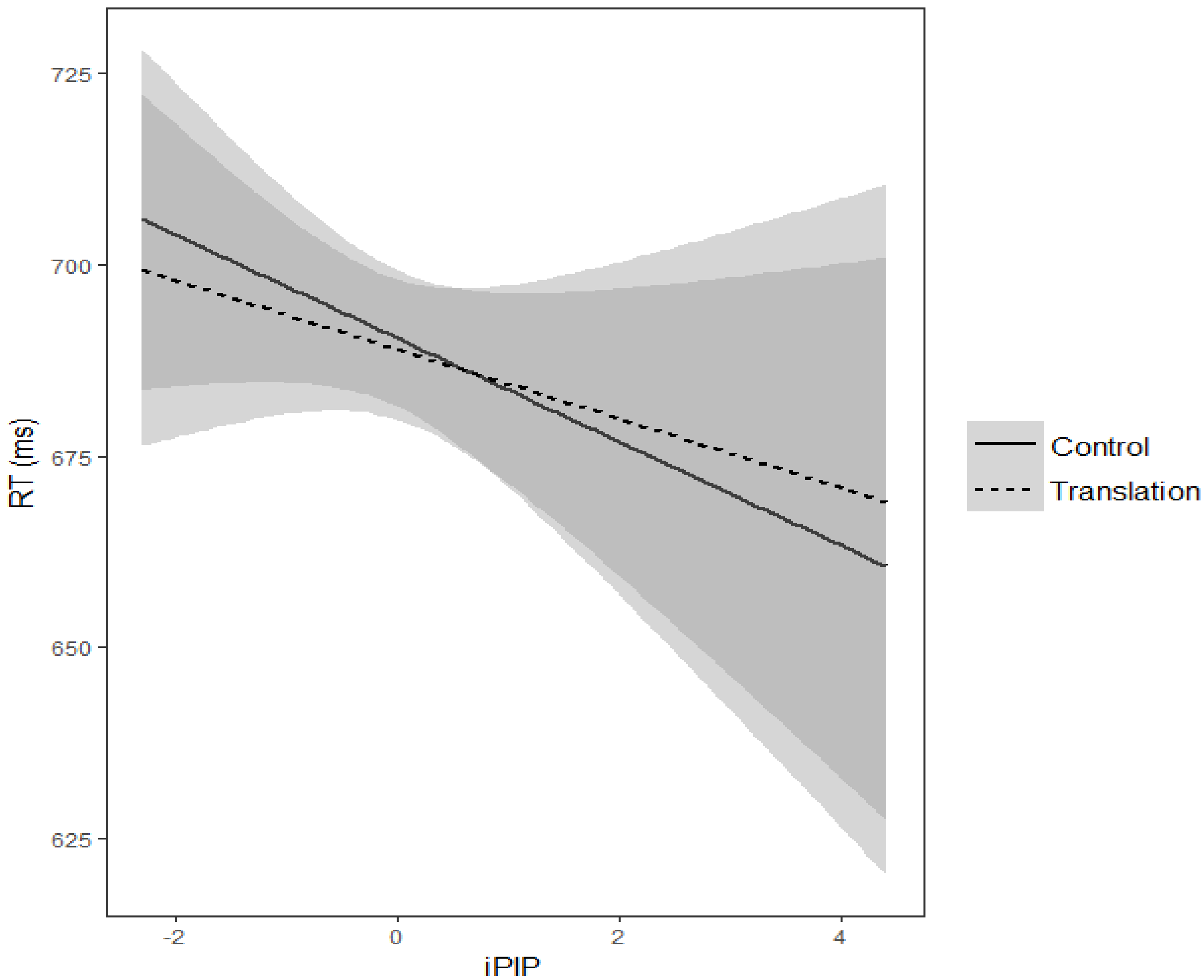

Figure 45. Response times as a function of prime and scaled iPIP, Experiment 4 New trials. Shaded areas represent 95% confidence intervals.

6.2.3.2.3 Prime x PIP Analysis

The additive model was once again favoured over the interactive model, *BF* = 48.71, θ = 2.07. Again, none of the effects were significant in any analysis, *t*s < 1.21, *p*s > .22. The effects of prime and PIP on the RTs of New trials are shown in Figure 46.

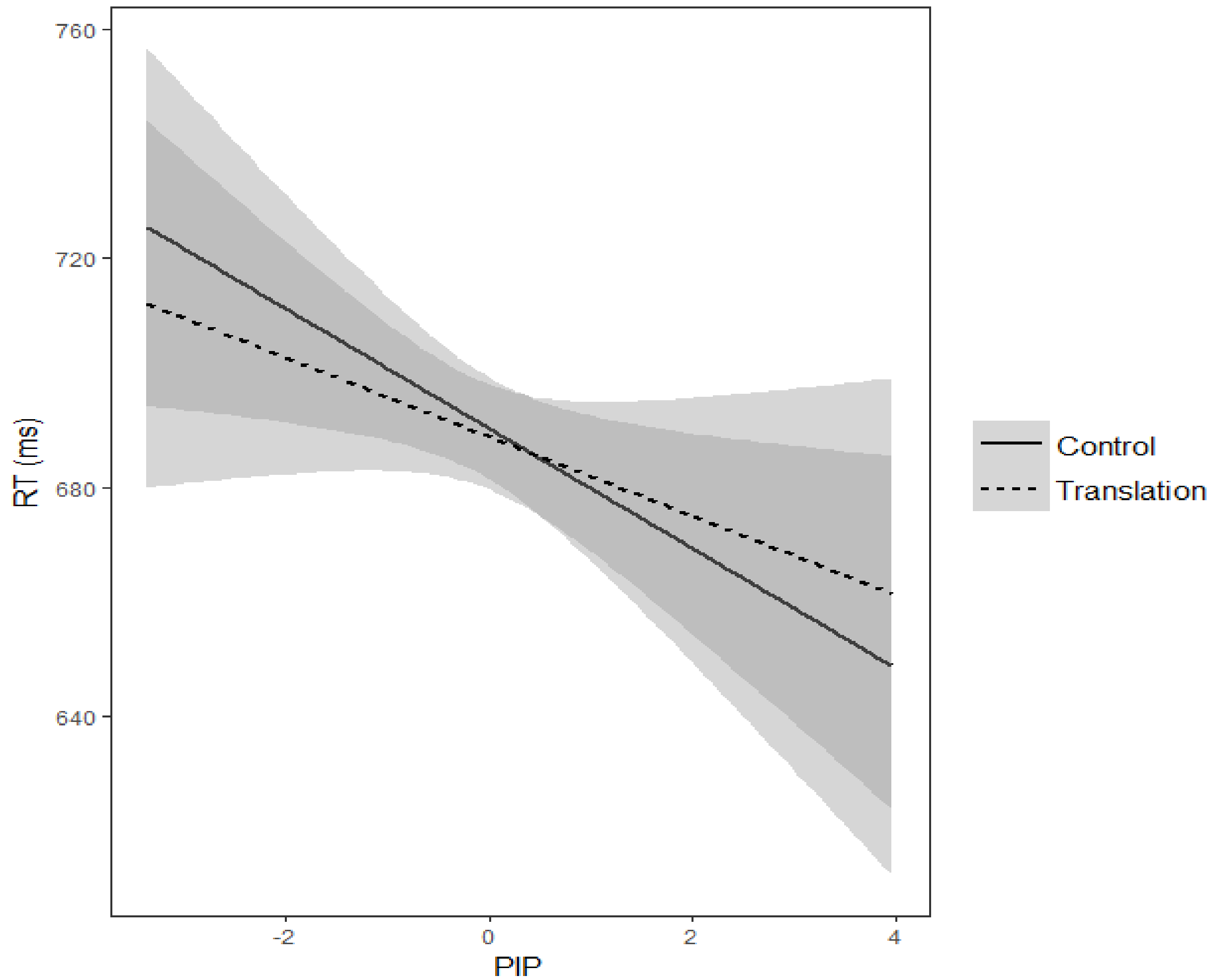

Figure 46. Response times as a function of prime and PIP, Experiment 4 New trials. Shaded areas represent 95% confidence intervals.

6.2.3.2.4 Prime x Order Analysis

None of the effects were significant in this analysis, *t*s < 1.38, *p*s > .16.

6.2.4 Reaction Time Analysis, Block 1 Only

6.2.4.1 Old Trials

Due to models being unable to converge when sPIP, iPIP, or PIP were included as fixed effects in any analysis, the effect of prime was assessed in the first block to test whether a priming effect was produced during the initial phase of the task, but then was lost in blocks 2 and 3. However, the effect of prime in the first block was nonsignificant, $t < 1$. Targets that were preceded by a translation prime ($M = 694$ ms) produced identical latencies to targets that were preceded by a control prime ($M = 695$ ms).

6.2.4.2 New Trials

Once again, there was no effect of prime in the first block for new trials, $t < 1$. Targets that were preceded by a control prime ($M = 694$ ms) produced identical response times to targets that were preceded by a translation prime ($M = 697$ ms).

6.2.5 Error Analysis, Full Data

6.2.5.1 Old Trial Analysis

6.2.5.1.1 Prime x sPIP Analysis

The additive model was favoured over the interactive model, $BF = 47.88$, $\theta = 1.93$. This model involved a nonsignificant effect of the prime on error rates, $z < 1$. This model involved a significant effect of sPIP on error rates, $\beta = 0.46$, $SE = 0.22$, $z(4565) = 2.10$, $p = .035$, which is shown in Figure 47, but the two-way interaction was nonsignificant, zs < 1. In particular, subjects in Tertile 1 ($M = 7.75\%$) produced significantly smaller error rates than subjects in Tertile 2 ($M = 21.76\%$) and Tertile 3 ($M = 20.51\%$).

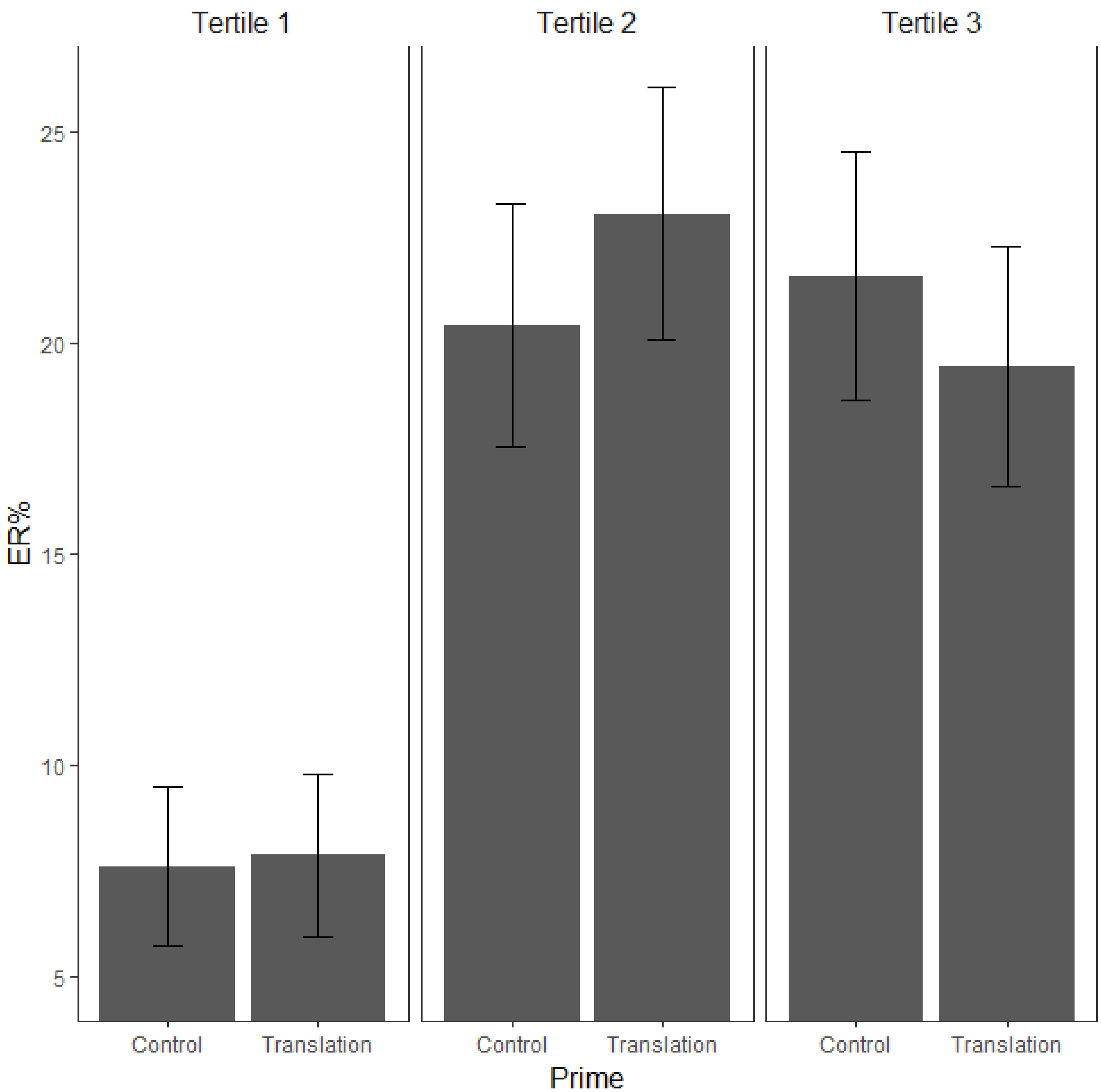

Figure 47. Mean error rates as a function of prime and sPIP tertile, Experiment 4 Old trials. Error bars represent 95% confidence intervals.

6.2.5.1.2 Prime x iPIP Analysis

The additive model was once again favoured over the interactive model, $BF = 62.42$, $\theta = 2.51$. None of the effects were significant in any analysis, $zs < 1.19$, $ps > .23$. The mean error rates for Old trials as a function of prime and iPIP tertile are shown in Figure 48.

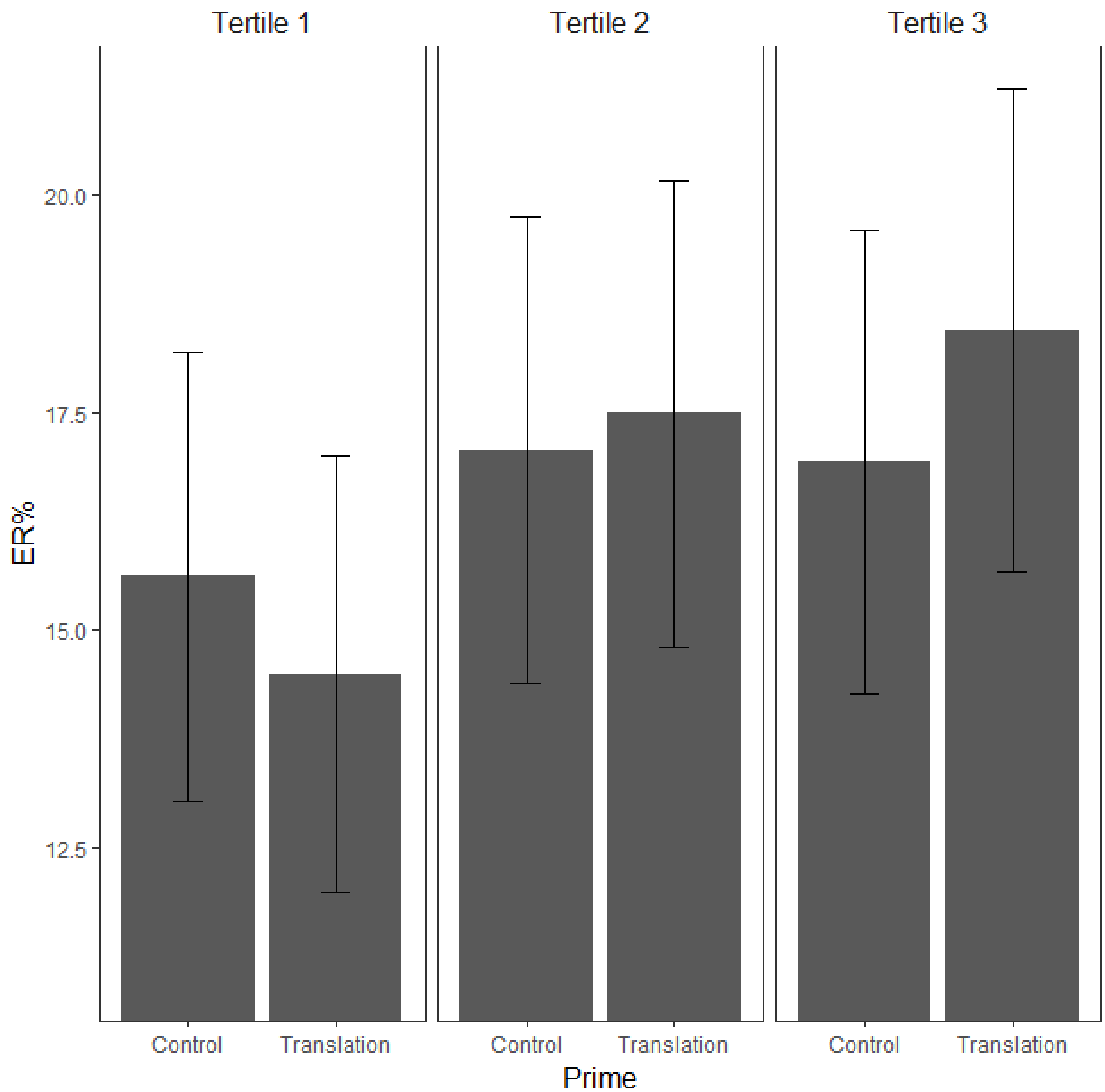

Figure 48. Mean response times as a function of prime and iPIP tertile, Experiment 4 Old trials. Error bars represent 95% confidence intervals.

6.2.5.1.3 Prime x PIP Analysis

The additive model was again favoured over the interactive model, *BF* = 64.95, θ = 2.61, which involved a significant effect of PIP on error rates, β = 0.16, *SE* = 0.07, *z*(4565) = 2.39, *p* = .017, as shown in Figure 49. Error rates in Tertile 1 (*M* = 9.26%) were lower than error rates in Tertile 2 (*M* = 20.25%) and Tertile 3 (*M* = 20.51%).

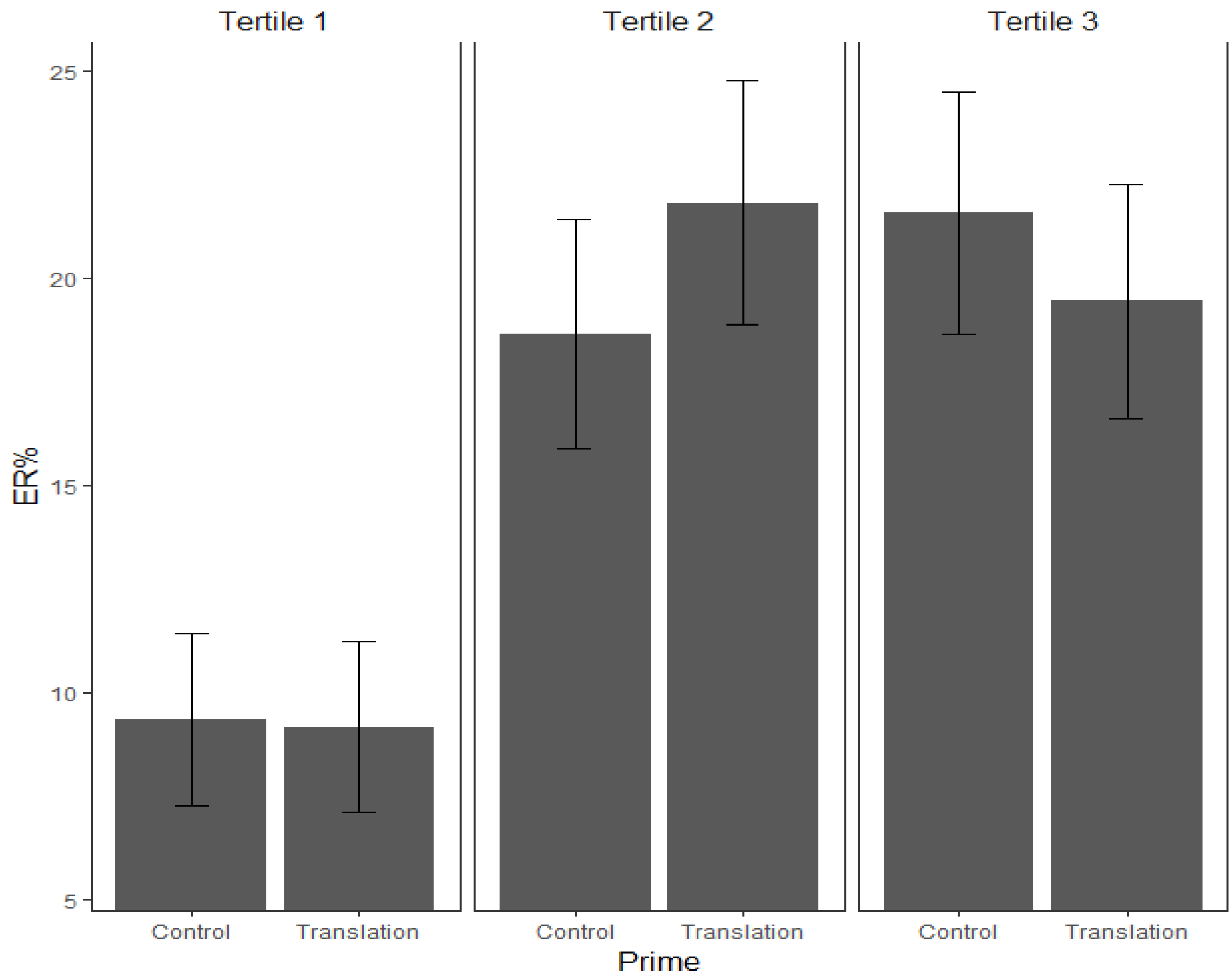

Figure 49. Mean error rates as a function of prime and PIP tertile, Experiment 4 Old trials. Error bars represent 95% confidence intervals.

6.2.5.2 New Trials Analysis

6.2.5.2.1 Prime x sPIP Analysis

For all analyses with New trials, the models would not converge unless random slopes were included. For the prime and sPIP analysis, sPIP was included as a random slope on items. The additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 1.81, but the interactive model was favoured in the relative likelihood analysis, θ = 13.72. When the effect of prime was excluded from a restricted model, this restricted model was favoured over the additive model in both analyses, *BF* = 22.02, θ = 22.02, indicating that the reason that the additive model was favoured over the interactive model was because the interactive model included the effect of prime, not because the interactive model included the interaction. This model did not involve a significant effect of prime, $z < 1.03$, $p > .30$. Targets that were preceded by translation primes ($M = 9.03\%$) and targets that were preceded by control primes ($M = 8.60\%$) produced comparable error rates. Although the effect of prime was nonsignificant, there was a marginally significant effect of sPIP, $\beta = 0.48$, $SE = 0.26$, $z(4561) = 1.81$, $p = .07$[34], and a significant two-way interaction between prime and sPIP, $\beta = .17$, $SE = 0..06$, $z(4561) = 2.66$, $p = .0079$, which is shown in Figure 50. Error rates were smaller in Tertile 1 ($M = 4.28\%$) than they were in either Tertile 2 ($M = 12.30\%$) or Tertile 3 ($M = 9.80\%$). While the effect of the prime was nonsignificant overall, the effect of the prime on error rates significantly differed between Tertile 1 (2.23% inhibitory effect), Tertile 2 (0.89% inhibitory effect), and Tertile 3 (1.93% facilitory effect).

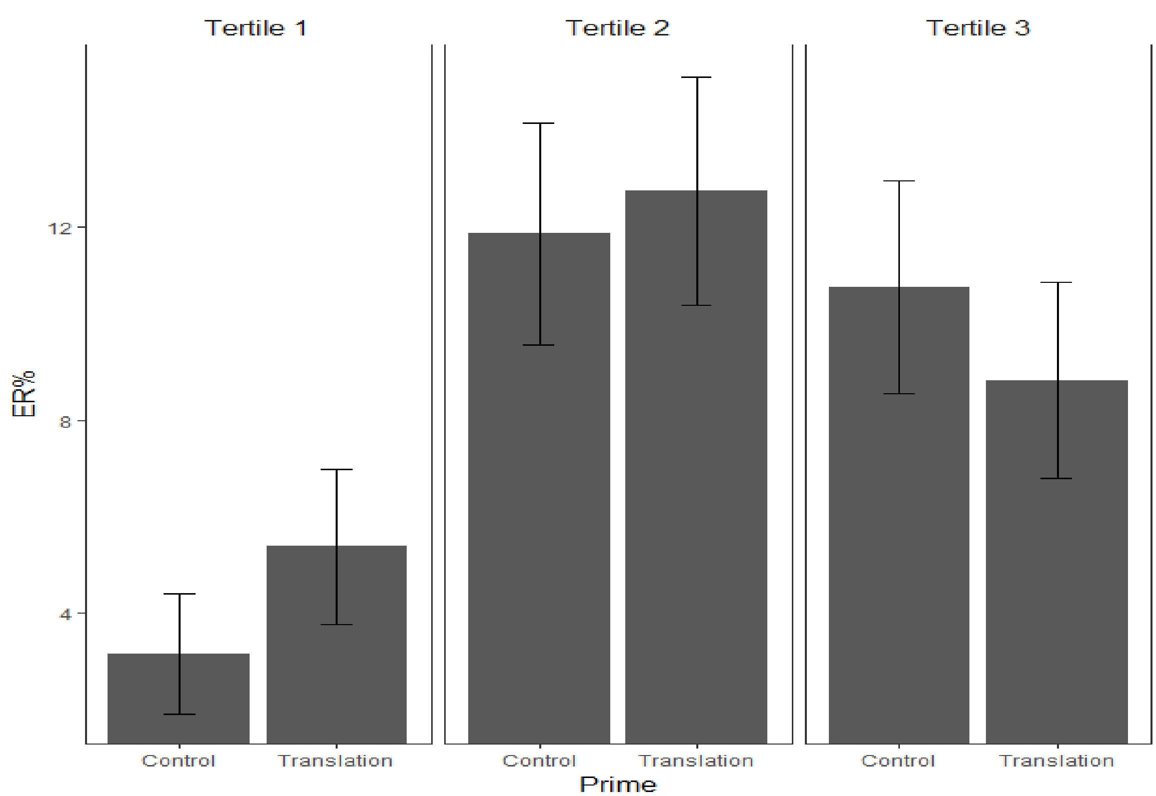

Figure 50. Mean error rates as a function of prime and sPIP tertile, Experiment 4 New trials. Error bars represent 95% confidence intervals.

[34] The effect of sPIP was significant when the screening criteria were loosened, $z(5030) = 2.74$, $p = .006$.

6.2.5.2.2 Prime x iPIP Analysis

For these analyses, iPIP was included as a random slope on subjects. The additive model was favoured over the interactive model, *BF* = 41.40, θ = 1.67. None of the effects were significant in this analysis, *z*s < 1.22, *p*s > .21[35]. The effects of prime and iPIP on the error rates of New trials are shown in Figure 51.

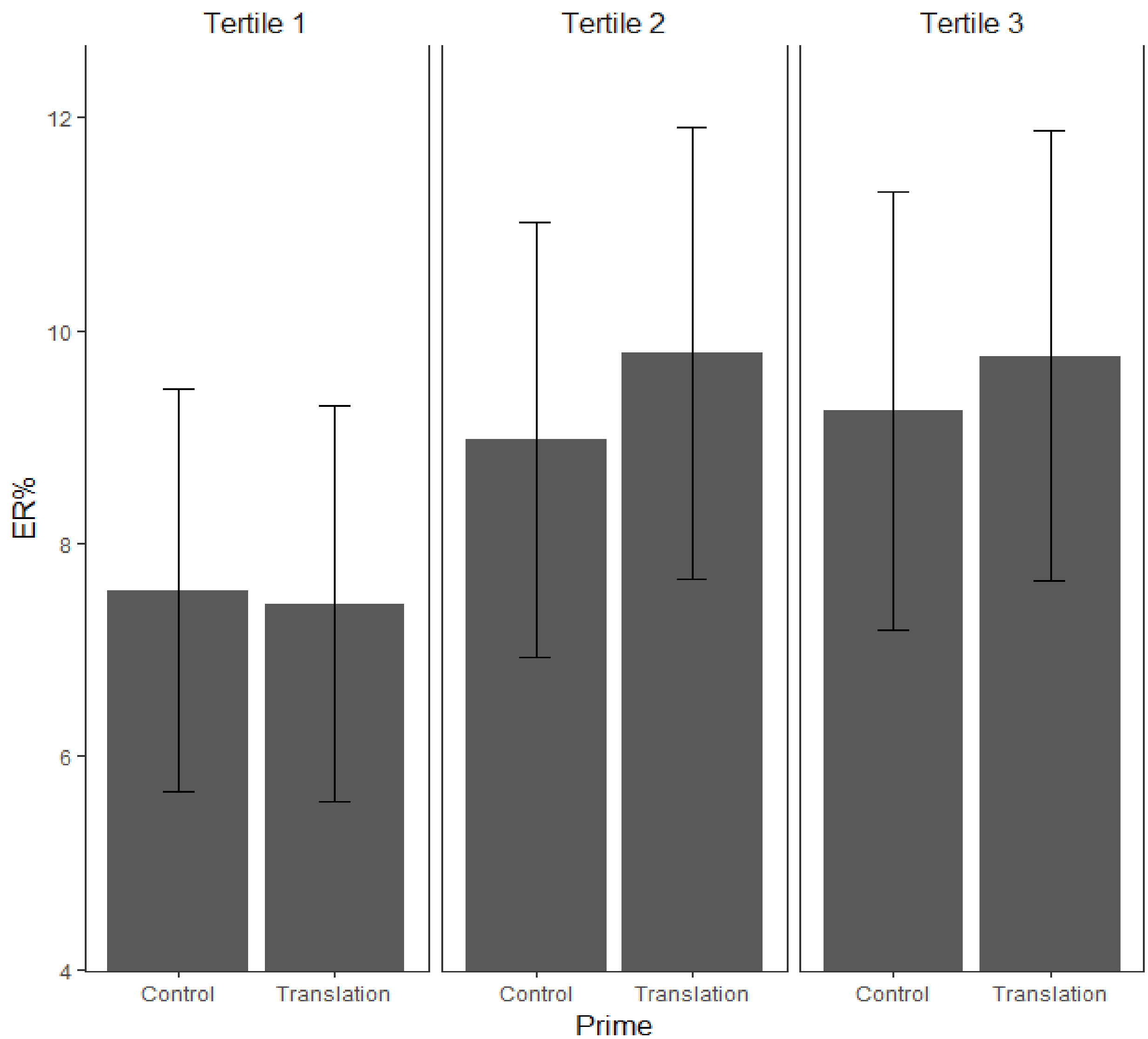

Figure 51. Mean error rates as a function of prime and iPIP tertile, Experiment 4 New trials. Error bars represent 95% confidence intervals.

[35] The effect of iPIP was marginally significant when the screening criteria were loosened, *z*(5030) = 1.73, *p* = .084.

6.2.5.2.3 Prime x PIP Analysis

The additive model was favoured over the interactive model, *BF* = 28.64, θ = 1.15, which involved a significant effect of PIP on error rates, $\beta = 0.65$, *SE* = 0.16, $z(4561) = 4.04$, $p < .0001$. Neither the effect of prime, $z < 1$, nor the two-way interaction were significant, $z < 1.32$, $p > .18$. The effects of prime and PIP on the error rates of New trials are shown in Figure 52. Errors in Tertile 1 (*M* = 4.67%) were smaller than errors in either Tertile 2 (*M* = 11.91%) or Tertile 3 (*M* = 9.80%).

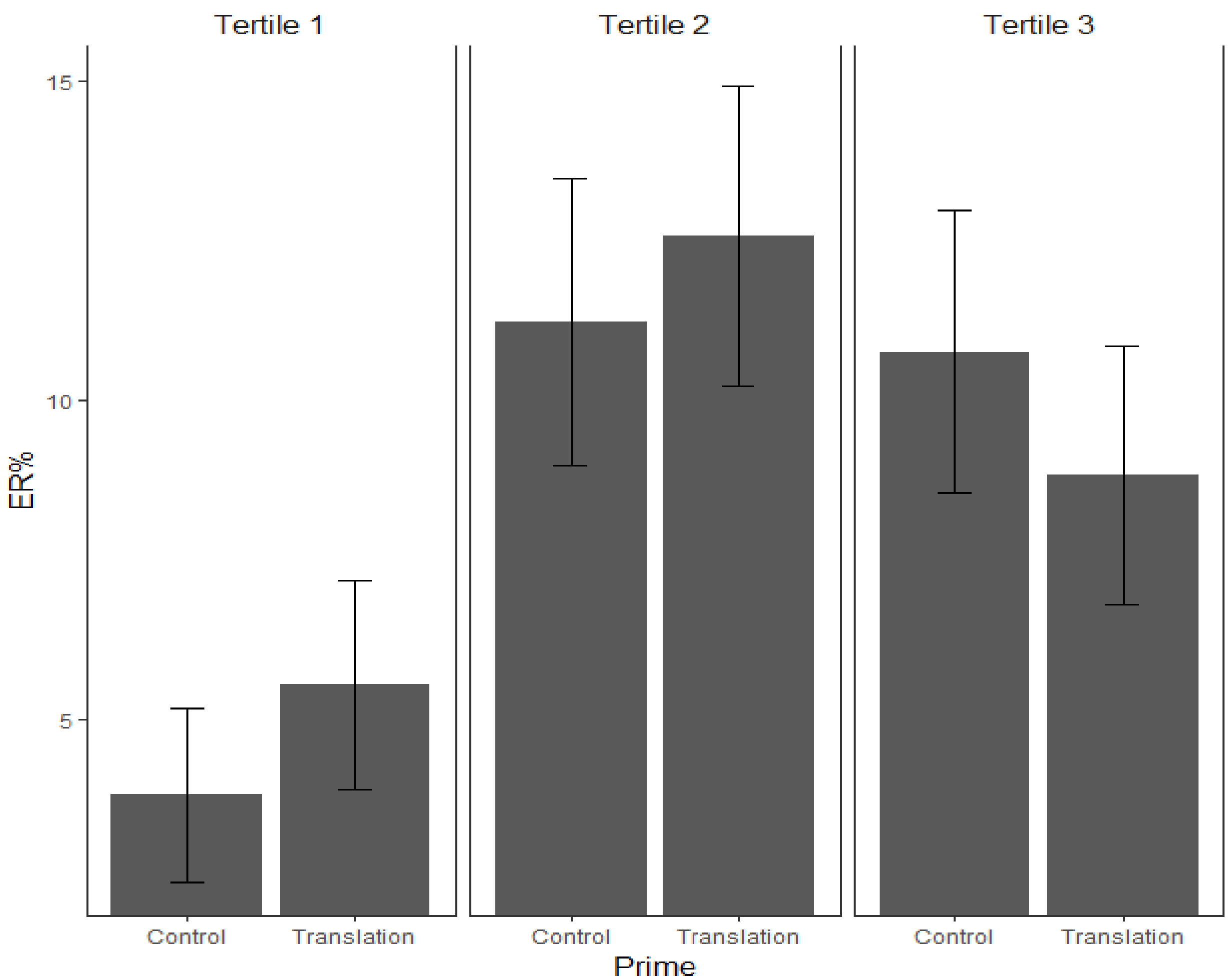

Figure 52. Mean error rates as a function of prime and PIP tertile, Experiment 4 New trials. Error bars represent 95% confidence intervals.

6.2.6 Error Analysis, Block 1 Only

6.2.6.1 Old Trials

The effect of prime was not significant in this analysis, $z < 1$. There was no significant difference in the error rates for targets preceded by translation primes (M = 16.09%) and control primes (M = 14.75%).

6.2.6.2 New Trials

The effect of prime was not significant in this analysis, $z < 1$. There was no difference in the error rates for targets preceded by translation primes (M = 7.83%) and for targets preceded by a control prime (M = 7.10%).

6.3 Discussion

Experiment 4 was conducted to test whether the assumptions of the Episodic L2 Hypothesis (e.g., Jiang & Forster, 2001) have some viability in terms of helping to understand the nature of bilingual language representations. In its present state, the Episodic L2 Hypothesis does not provide any theoretical mechanism that can explain why tasks that are assumed to rely on lexical and semantic processing would be sensitive to factors such as L2 proficiency, or sensitive to factors that presumably have a lexical locus of their effect, such as the frequency of L2 primes in lexical decision. One possible mechanism that could help to integrate the findings of Experiments 1-3 into the framework of the Episodic L2 Hypothesis would be to assume that L2 representations are initially episodic, but the locus of representation in memory changes over the time course of L2 acquisition, as learners become more familiarized with the language, and processing in L2 becomes more automatized. The transition away from episodic representations occurs as learners become highly familiarized with their L2, and acquire a deeper and broader level of understanding of words in their L2, and could occur at a faster rate for words that learners encounter more frequently in their use of L2. It was predicted, then, that if representations for words migrate from episodic to lexical memory, that priming effects in episodic recognition should be inversely related to learner-level factors such as L2 proficiency, age of initial acquisition, and the time that the subject has spent learning the L2, and word-level factors such as word frequency.

These predictions were not supported by the data. First, there was no overall effect of prime on RTs in Experiment 4, contrary to prior studies (e.g., Jiang & Forster, 2001; Witzel & Forster, 2012). The null effect of prime could not be attributed to fatigue effects, as the priming effect was null even when only the first block of data was analyzed, nor was there any difference in the priming effect when subjects completed Experiment 4 before Experiment 3 than when subjects completed Experiment 3 before Experiment 4. Second, many of the factors that predicted larger priming effects were contrary to these predictions. Subjects who reported higher global proficiency in English, who reported using English more often in other social environments outside of school and at home, and who reported learning English for a longer period of time and at a younger age tended to be more prone to producing facilitative priming effects in episodic recognition than subjects who were less proficient in English, reported using English less in daily life, and who reported learning English later in life. This trend was specific to Old trials, as there was no systematic relationship between the sPIP, iPIP, or PIP coefficients and priming effects in New trials. What these data suggest, instead, is that L2-L1 translation priming in episodic recognition is also facilitated by subjects' proficiency in their L2, much as it is in lexical decision and semantic categorization. A more complete overview of how these results could be accounted for is provided in the General Discussion.

Chapter 7

7 General Discussion

The present research was an attempt to examine L2-L1 masked translation priming effects under the assumption that it is a task-specific process and to understand what skills and linguistic behaviours were predictive of priming in each task. In part, the purpose of examining what skills and linguistic behaviours predicted translation priming across tasks was to test whether the results in these tasks can be accommodated by current theories of bilingual memory, such as the BIA+ (Dijkstra & van Heuven, 2002), the RHM (Kroll & Stewart, 1994), the Sense Model (Finkbeiner et al., 2004), and the Episodic L2 Hypothesis (Jiang & Forster, 2001). However, another reason for examining the skills and behaviours that predict priming was to understand why a dissociation has occurred between lexical decision, semantic categorization, and episodic recognition in general and, in particular, why translation priming effects arise consistently in semantic categorization tasks (e.g., Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), but not in the lexical decision task (e.g., Gollan et al., 1997).

The present research has produced several insights. First, all three tasks showed an interaction between prime and proficiency, as measured by the sPIP score. Subjects to whom could be attributed higher sPIP scores tended to produce larger priming effects than subjects to whom could be attributed lower sPIP scores in each task with these scores being largely computed on the basis of subjects' competency with their L2 across different domains, and their use of their L2 in daily life. Finding that the sPIP score interacted with priming, then, provides good evidence that the priming effect is sensitive to L2 proficiency. Further, the results have shown that priming effects are also sensitive to item-specific factors, specific to both the prime and the target, as measured by the iPIP score. With the exception of Experiment 4, items to which could be attributed higher iPIP scores also tended to produce larger priming effects than items to which could be attributed lower iPIP scores.

Second, there appears to be a dissociation between the skills, behaviours, and item-specific factors that predict L2-L1 priming across different tasks. In lexical decision, rather than any objective, standardized measure of English proficiency, the largest subject-based predictors were subjects' self-rated listening and writing abilities in English, and the self-rated reading and listening abilities in Chinese, while the largest item-based predictor was the CELEX frequency of the English prime. Subjects who reported having better spoken comprehension abilities in English, and better expressive writing abilities in English produced larger priming effects than subjects who reported being weaker in these domains. Targets that were primed by high-frequency translation primes produced larger priming effects than targets that were primed by low-frequency translation primes.

In semantic categorization, the largest subject-based predictor was the amount of time English was used by subjects across different social contexts, specifically, the use of English at school, and in other social contexts. The largest item-based predictor was the Chinese target's frequency. Subjects who reported using their L2 more in day-to-day life across a wider range of social contexts produced a larger priming effect in the semantic categorization task than subjects who used their L1 more heavily outside of the home, and high-frequency exemplar targets produced larger priming effects than low-frequency targets. There was also an effect of prime typicality. Targets with translation equivalents that are more typical members of the target category tended to produce larger priming effects than targets that had atypical translation equivalents and, hence, were more likely atypical themselves). Finally, in the speeded episodic recognition task, the largest predictors of priming were self-rated writing, reading, speaking, and listening proficiency in English, the number of years subjects had been learning English, and self-rated writing and listening proficiency in Chinese. Subjects who reported being more proficient in English produced larger priming effects. The implications of these findings are discussed below.

7.1 Translation Priming In Lexical Decision

With respect to the lexical decision task, these results contribute to a mounting body of recent evidence that priming in the lexical decision task is related to subjects' competency in their L2 (e.g., Nakayama et al., 2016). These results also provide the first evidence that masked translation priming effects in lexical decision are sensitive to individual differences in specific domains of L2 knowledge and proficiency, rather than global proficiency levels. Specifically, these results show that translation priming in lexical decision depends on subjects' writing abilities in English, and is negatively associated with subjects' reading and writing abilities in Chinese.

These results also provide some of the first evidence that masked translation priming effects are sensitive to the frequency of both the prime and the target. Priming effects were larger for targets that were preceded by high-frequency translation primes than targets that were preceded by low-frequency translation primes, and priming effects were larger when the target was low-frequency than when the target was high-frequency. These results are very similar to the results of Nakayama et al.'s (2012, 2013) studies, which found that L1-L2 priming effects are larger when the subjects are less proficient in their target language. Experiment 1's results suggest that this pattern is also true in the L2-L1 direction, when subjects are less proficient in their L1. These results additionally show that L2-L1 masked translation priming in lexical decision is sensitive to the frequency of both the prime and the target. Priming effects were larger when the frequency of the target was lower, and the frequency of the prime was higher. Again, these results bear similarities to the results of Nakayama et al.'s studies, which found the same effect of target frequency. Overall, such results are consistent with the notion that the facilitation associated with translation priming in lexical decision is dependent on the difficulty associated with the processing of targets. The more difficult it is for subjects to process the targets, the more influence a prime can exert in driving decisions in the task.

Models such as the Sense Model (Finkbeiner et al., 2004), which assume that the priming asymmetry in lexical decision is due to asymmetries in the semantic representations of L1 and L2 words, would require several assumptions to account for these findings. With respect to the findings with sPIP, the Sense Model would have to assume that, as bilinguals become more proficient in their L2, the L2 senses that bilinguals acquire are largely shared with their L1 translation equivalent, and that the acquisition of these overlapping senses would be sufficient to produce facilitative effects. Only senses that are shared across languages would contribute to larger priming effects, as the acquisition of L2-specific senses would have no impact. With respect to iPIP, the Sense Model would have to account for why priming effects were also influenced by the frequency of the prime and target. It could be argued that the number of senses associated with words is correlated with word frequency, and argue that the effect of prime frequency observed in the iPIP score was actually due to the primes having more senses[36], but it

[36] There was a weak positive correlation between number of senses and prime frequency, $r(98) = .18$, $p < .08$, $R^2 = .031$.

would have to be again assumed that these senses tend to be shared with the L1 translation equivalent, at least across a sufficient enough number of the items to produce facilitation. The likelihood of both of these assumptions being met, however, is questionable, as these assumptions would require the systematic increase in overlap between L2 and L1 senses across words and subjects, when many of the Chinese words included in Experiment 1 had very few senses (e.g., 法案 refers unambiguously to a legislative bill), or had senses that do not overlap with their L2 translation equivalent (e.g., 玻璃 can refer to either glass, or any film-like material that possesses the same transparency as glass, such as cellophane, nylon, or plastic). It would thus be more parsimonious to argue that these results are consistent with the priming asymmetry effect being driven by factors such as bilinguals' productive abilities with L2 written text, reading and writing abilities in their L1, and the frequency of occurrence of primes and targets.

With respect to the Episodic L2 Hypothesis (Jiang & Forster, 2001), Experiment 1's results cannot be accommodated by this account in its present state, as that model predicts that no priming effects should occur in lexical decision, and does not presently make any assumptions about whether L2 representations change from being stored in episodic memory to lexical memory over the course of L2 acquisition. However, that is not to say that the model cannot be augmented to account for some of these findings. An alternative framework, which could provide at least a partial account, is discussed in greater detail below, when discussing the results of the speeded episodic recognition task.

While there is no clear mechanism for how these results could be accommodated by the RHM (Kroll & Stewart, 1994), such a pattern of findings can be accommodated by the BIA+ (Dijkstra & van Heuven, 2002) if it is assumed that proficiency in the domain of writing in one's L2 impacts resting-level activity of L2 representations in the word identification subsystem of the model. One possible locus of writing proficiency could be within the lexical orthographic layer in the model. Bilinguals who are highly skilled and familiarized with the orthographic system of their L2 would be predicted to have higher resting-level activity in this domain than bilinguals who have less skill and familiarity with their L2's orthographic system. When a prime is presented for a very brief period of time, the sublexical orthographic representations become activated, which, in turn, send activity to orthographic lexical units. Finally, the orthographic lexical layer sends activity to units in the semantic layer, and the task/decision subsystem then uses activity from the word identification subsystem to make a task-appropriate decision. For bilinguals who are highly skilled and familiarized with their L2's orthographic system, the resting-level activity within the lexical orthographic layer has a head start, and there is less of a temporal delay in the activation of L2 orthographic representations, allowing the prime to successfully preactivate the representations of the target, resulting in a priming effect. On the other hand, bilinguals who are less skilled and familiarized with their L2's orthography would show a temporal delay in the activation of L2 representations. As such, masked primes are less likely to preactivate the representations associated with the target, and no priming effect is observed. It should be noted that, at least for lexical decision, this account would appear to predict that resting-level activity within lexical orthography is more affected by bilinguals' productive abilities in their L2 writing system, rather than their receptive abilities, a notion which would be consistent with Swain's (1985, 2000) output hypothesis.

The BIA+ Model (Dijkstra & van Heuven, 2002) can additionally account for the effects of prime and target frequency on translation priming by assuming that the frequency of occurrence of the prime and target affects the general resting-level activation of the representations of each word. High-frequency primes would have higher resting-level activity than low-frequency primes. As a result, there is less of a delay in the activation of the word's representation, and the semantics of the prime are more consistently accessed as a result. Likewise, the resting-level activation of lower-frequency targets would be lower, meaning that the activation of lexical and semantic representations associated with the words would be slower. Under circumstances where the prime is high-frequency, and the target is lower-frequency, the resting-level activity of the prime and target is more similar, and there is a greater opportunity for the prime to facilitate the processing of the target by preactivating the relevant semantic representations associated with the target.

Beyond the Sense Model, (Finkbeiner et al., 2004), the RHM, (Kroll & Stewart, 1994), and the BIA+ model, (Dijkstra & van Heuven, 2002), there is one other set of principles which may help explain the results of Experiment 1. The findings of Experiment 1 are largely consistent with the notion that the ability to process the prime in an efficient manner is dependent on the integrity and quality of orthographic lexical representations in a bilingual's L2. This interpretation is consistent with findings from other studies that show that variations in exposure to print affect behavioural results across a number of domains, including lexical decision latencies (e.g., Chateau & Jared, 2000), repetition priming effects (e.g., Lowder & Gordon, 2017), gaze durations on words in eye-tracking (e.g., Gordon, Lowder, & Hoedemaker, 2016; Moore & Gordon, 2015, 2016; Taylor & Perfetti, 2016), spelling ability (e.g., Stanovich & West, 1989), verbal fluency (e.g., Stanovich & Cunningham, 1992), vocabulary knowledge (e.g., Stanovich, West, & Harrison, 1995; West & Stanovich, 1991; Mol & Bus, 2011) and reading comprehension (e.g., Martin-Chang & Gould, 2008; Mol & Bus, 2011).

Such results have often been accounted for within the framework of Perfetti and colleagues' (Perfetti, 1985, 2007; Perfetti & Adlof, 2012; Perfetti & Hart, 2002; see also Yap, Tse, & Balota, 2009) Lexical Quality Hypothesis. This account is based on the idea that reading skills such as comprehension, are affected by what those authors refer to as the "lexical quality" of word representations. Perfetti (2007) argues that efficient reading processes are underpinned by two major components of knowledge: 1) knowledge about word forms, which includes grammatical knowledge as well as knowledge of spelling and pronunciation, and 2) knowledge of word meanings. Perfetti used two criteria to define the "quality" of lexical representations: precision and flexibility. A lexical representation is precise to the extent that the mapping between the form and meaning components of word knowledge is highly stable, and "facilitates activation of the lexical representation corresponding to the sensory input and minimizes activation of competing alternatives" (Andrews & Hersch, 2010, p. 312). The flexibility of a word representation refers to the knowledge of the range of meanings that a word can take on, independent of context. Precision and flexibility are both required for the efficient retrieval of a word's identity. Precision is required, for example, when discriminating between words such as *potion* and *option*, or *would* and *wood*, which may be spelled or pronounced similarly, but are different words. Flexibility is required, for example, because words such as *subject* can mean "a person that is being discussed, described, or dealt with", "a branch of knowledge studied or taught in a school, college, or university", or "cause or force to undergo (a particular experience or form of treatment)", and to understand the use of *subject* in everyday use, one must understand the range of these meanings. Finally, both precision and flexibility are required when pronouncing *desert* in sentences such as "they intended to *desert* the man in the *dessert*". The quality of the lexical representations is determined by the combination of these two factors.

Perfetti (2007) further argues that lexical quality can manifest in orthography, phonology, grammar, meaning, and in the extent to which orthographic, phonological, and semantic components are bound together. A high-quality orthographic lexicon would be one in which the orthographic system is fully specified, in that the letters that compose the orthographic representations within this system are held constant, and these representations remain stable over time. In phonology, a high-quality representation would be one in which phonology is word-specific, and grapheme-phoneme correspondences are sensitive to context (e.g., the difference between the pronunciation of *record* in "I broke my personal *record*" and "I want to *record* a new song"). In grammar, a high-quality representation would be one in which all of the grammatical classes and morphosyntactic inflections are properly represented. In meaning, high-quality representations are ones in which the meaning is not bound by context, and the range of meaning dimensions is specified to the point that one can discriminate between words that are semantically similar. Finally, a high-quality lexical representation would be one in which the orthographic, phonological, morphosyntactic and semantic components are bound together tightly. The quality of these lexical representations is assumed to have processing consequences during reading, as it affects the stability and reliability with which word identity is retrieved from an orthographic or phonological input, the synchronicity with which the components of a lexical representation are activated and retrieved as a coherent word identity, and the ability to integrate the meaning of words into one's comprehension of what is being read. The crux of Perfetti's theory is that greater practice and experience with these components of knowledge leads to efficient, rapid retrieval of word identity.

While much of the present work has been aimed at investigating the impact of exposure to one's L2 orthography on cognitive processes, an account of this sort can certainly be extended to allow an understanding of the effects of experience bilinguals get by actively using their L2 orthography, as research has also shown that factors associated with writing ability, such as spelling, are also associated with better phonological processing skills (e.g., Allyn & Burt, 1998; Pennington, Lefly, Van Orden, Bookman, & Smith, 1987), and better visual word identification abilities (e.g., Burt & Fury, 2000; Burt & Tate, 2002). Much like being exposed to print, actively using one's L2 to formulate ideas in print can gradually improve the specification, the precision, and the flexibility of L2 lexical representations. In the context of a masked translation priming task, the improved precision of L2 lexical representations leads to more efficient and reliable retrieval of the prime's meaning. In a task that stresses lexical processing such as lexical decision, then, the information that is most salient to the task would be how specified the orthographic lexicon is, and how well the semantic and orthographic components have been bound together within the lexical representation. It is assumed, then, that more experience actively using one's L2 in expressive writing improves the precision and flexibility of the L2 orthographic lexical system, and strengthens the binding between the L2 forms and meaning. An additional factor that is assumed to affect the binding between L2 form and meaning is the frequency of the L2 word. Higher frequency L2 words are ones which L2 learners encounter more often throughout daily life, and, as a result, the binding between form and meaning is tighter than for low-frequency words.

There are, however, a few caveats. First, none of these interpretations appears to have a way of addressing the fact that the largest predictive factor associated with L2 competency was the comprehension of spoken English. Such a result need not be surprising, however. Even if listening and writing represent knowledge of language in different modalities, it is well-known that skills in spoken language play a major role in the development of reading and writing skills (e.g., Bishop & Snowling, 2004; Gillam & Johnston, 1992; Kroll, 1981; McCutchen, 1986) Research has shown that receptive abilities develop earlier in the course of language development than expressive abilities (e.g., Guess, 1969; Huttenlocher, 1974). From a developmental perspective, the first skill that one typically acquires in language development is the comprehension of spoken language. Regardless of whether the language is learned from birth, as would be the case with one's L1, or whether one is acquiring the language at a later stage of life, the acquisition of passive knowledge of different grammatical structures, vocabulary, pragmatic understanding of language use, and understanding of word meanings that would be associated with spoken comprehension is an essential prerequisite for effectively acquiring other abilities in a language. Skills such as reading, writing, and speaking would not develop if this knowledge didn't exist to support the acquisition of these skills (e.g., Dockrell & Connelly, 2009).

Second, these results also imply that L2-L1 translation priming in lexical decision is also affected by subjects' reading and writing abilities in their L1. While these results suggest that subjects who are weaker in productive and receptive orthographic tasks in their L1 are more prone to utilizing the L2 prime to drive decisions on targets, one unanswered question that these results raise is whether these subjects became weaker readers and writers in Chinese as a consequence of becoming better readers and writers in English, whether they were always poor readers and writers in Chinese prior to acquiring English, or whether they lagged behind other subjects because they were more prone to dividing their frequency-of-use of each language, resulting in weaker reading and writing skills in their native language (e.g., Gollan, Montoya, Cera, & Sandoval, 2008). This distinction is important, as it has implications for understanding the consequences that learning a second language has on processing in an L1. If subjects did not become weaker readers and writers in Chinese as a consequence of becoming better readers and writers in English, that would imply that learning how to read and write in English had no consequences for processing in their L1, and that these participants were more predisposed to benefitting from the prime due to having had weaker reading and writing abilities in Chinese prior to learning English. The latter idea would imply that becoming more proficient in an L2 has had consequences for subjects' processing abilities in their L1, and this combination of becoming more proficient in an L2 while one's L1 skills deteriorate is what resulted in subjects showing a larger impact of the L2 prime on the L1 target.

Regarding the latter possibility, this idea is not one that is new. Research looking at the effects of L1 processing on L2 acquisition is quite extensive, with research showing evidence of a negative transfer when the bilinguals' two languages use different writing systems (e.g., Bialystok, 1997; Holm & Dodd, 1996; Liow & Poon, 1998; however, see Wang, Perfetti, & Liu, 2005), and showing a negative relationship between the breadth of vocabulary knowledge in L1 to the breadth of vocabulary knowledge in L2 (e.g., Ordonez, Carlo, Snow, & Mclaughlin, 2002). More recently, Kaushanskaya, Yoo, and Marian (2011) examined the effects of second-language exposure on vocabulary and reading skills in subjects' native language. Kaushanskaya et al. compared English-Spanish and English-Mandarin bilinguals, who were tested on vocabulary knowledge and reading fluency in English, and subjects provided additional information about their history of L2 acquisition, including the age at which the language was acquired, the amount of exposure to the L2, L2 proficiency, and preference of L2 use. Kaushanskaya et al. found evidence that processing in an L2 can not only influence processing in subjects' L1, but that the manner in which processing in an L2 influences L1 processing is influenced by the extent that the two languages are similar. For the English-Spanish bilinguals, Kaushanskaya et al. found that reading proficiency in Spanish was positively associated with reading proficiency in English. Critically, for English-Mandarin bilinguals, self-reported Mandarin proficiency was negatively associated with English reading proficiency. These results suggest that L1 writing and reading skills are impacted by the degree of typological overlap between the two languages. These results show that subjects who have weaker abilities in their L1 in reading and writing, but have relatively strong expressive abilities in L2 writing benefit more from translation priming than subjects who are strong readers and writers in their L1, and weaker writers in their L2.

A final caveat worth noting is that self-rated L2 writing abilities may reflect a wide variety of different processes and skills, from orthographically based factors such as spelling and orthographic coding efficiency, to the broader knowledge of the nuances of the language that one is communicating in that allows one to effectively formulate meaningful, precise, and grammatically-correct expressions in that the language. Certainly, in the literature on writing fluency, the components of how to define writing fluency have not been universally agreed upon. Whereas some researchers define writing fluency as the ability to produce written language quickly, appropriately, and coherently (e.g., Wolfe-Quintero, Inagaki, & Kim, 1998), others base their definition on the rate of text composition (e.g., Sasaki, 2000), the quantity of text produced (Baba, 2009), the speed which with writers retrieve lexical representations while writing (Snellings, van Gelderen, & de Glopper, 2004), and some use a variety of other criteria to assess writing ability (see Abdel Latif, 2013, for a full review). As such, while a number of the interpretations and possible explanations offered in this dissertation have focused on orthographic coding efficiency, and the quality of L2 lexical representations, the best measure of writing skill may reflect a wide array of other factors. The task of identifying how these specific components of L2 writing ability contribute to cross-language translation priming is one that will be a subject of future research.

One avenue for future research is in examining the effects of orthographic awareness and orthographic decoding efficiency on L2-L1 translation priming. Studies that examine individual differences in L2 spelling abilities, orthographic lexical precision, and knowledge of word forms, for example, could provide valuable information on the role of orthographic knowledge in mediating semantic access in L2-L1 priming in alphabetic languages, and would provide insight into how such knowledge contributes to the acquisition of reading skill.

If the results of Experiment 1 are any indication, perhaps the most valuable avenues for future research lie in studying the effects of vocabulary knowledge on cross-language lexical decision performance. Research has shown that vocabulary knowledge is one of the most powerful predictors of early writing, speaking, and reading abilities in children between the ages of 8-16 (e.g., Dockrell & Connelly, 2009). More importantly, in a task such as lexical decision, the usefulness of primes would be dependent on the knowledge that one has about words in the priming language. The role of vocabulary knowledge could be particularly important, for example, when the prime-target relationship is purely semantic in nature, as when there is no orthographic or phonological overlap that could aid in the decision process, and that knowledge of L2 vocabulary could be essential in extracting the semantics from the prime to preactivate the target. And yet, very few studies have examined the role of vocabulary knowledge in language processing.

Research that has been done on vocabulary knowledge, however, suggests that vocabulary knowledge has a significant impact on tasks such as lexical decision (Yap, Balota, Tse, & Besner, 2008), naming (e.g., Bialystok, Craik, & Luk, 2008), reading (e.g., Federmeier, McLennan, De Ochoa, & Kutas, 2002), speech perception (Banks, Gowen, Munro, & Adank, 2015), speech production (e.g., Rodriguez-Aranda & Jakobsen, 2011), and L2 writing production abilities (e.g., Coxhead, 2007, 2018; Johnson et al., 2016; Laufer & Nation, 1995; Staehr, 2008; Zhong, 2016). In monolingual studies, vocabulary knowledge has also been shown to interact with factors such as word frequency (e.g., Mainz, Shao, Brysbaert, & Meyer, 2017), and how factors such as word frequency statistically combine with other factors, such as semantic priming (e.g., Yap, Tse, & Balota, 2009). Although such results have shown that vocabulary knowledge typically reduces the effects of factors such as word frequency in lexical decision in monolingual task contexts (e.g., Brysbaert, Lagrou, & Stevens, 2017; Mainz et al., 2017; Monaghan et al., 2017; Yap et al., 2009), there is little reason to believe that such a trend would also occur in cross-language tasks such as translation priming, specifically if the factor of interest is the knowledge of the priming language vocabulary. Under those circumstances, having larger, well-specified vocabularies should increase priming effects.

If anything, one contributing factor to the asymmetry between L1-L2 and L2-L1 tasks is the discrepancy between vocabulary knowledge in L1 and L2, as bilinguals’ L2s usually have sparser vocabulary and less well-defined lexical representations compared to their L1s. When L2 words are used as targets, having primes from L1 thus produce a benefit because there is a greater opportunity for the prime to aid the lexical processing of the target. When L1 words are used as targets and L2 words are used as primes, as was the case here, however, the unstable representations of the primes, coupled with the sparser vocabulary in L2, means that there is a reduced likelihood that the prime will aid in the lexical processing of the target, and there is a lower likelihood that the prime is even a familiar part of the subject's vocabulary. A further discussion of the role of vocabulary knowledge in lexical decision and semantic categorization is found below.

7.2 Translation Priming in Semantic Categorization

The results of the semantic categorization tasks have several implications. First, these results demonstrated that, much like the lexical decision task, there are sets of factors that predict the likelihood that subjects can access the semantics of the prime in a way that affects decisions on the target. Consistent with past research (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), these results showed that while the semantic categorization task did produce a larger priming effect than the lexical decision task, replicating prior research, that the magnitude of priming effects systematically varied with proficiency, as measured by the amount of time subjects used their L2 across a variety of social contexts, and their self-rated L2 verbal productive abilities. Further, subjects who tended to rate themselves as having weak verbal productive abilities, and who used their L2 more sparsely in daily living tended to produce weaker, or even null priming effects. Finally, unlike lexical decision, priming in the semantic categorization task was facilitated by the target frequency, rather than the prime frequency, suggesting that the processes that drive translation priming in semantic categorization and lexical decision are qualitatively different.

Once again, these semantic categorization results are difficult to reconcile with the Sense Model (Finkbeiner et al., 2004) in its current form, as the Sense Model assumes that L2-L1 priming in a semantic categorization task is not contingent on the proportion of primed-to-unprimed senses, but by whether the L2 prime activates senses that denote category membership. For most translation equivalents, bilinguals would usually learn the senses associated with L2 words that would contain such information first. Such senses should be acquired by even less proficient bilinguals, and such bilinguals should produce significant priming effects in this task. As such, the Sense Model would have trouble accounting for why priming effects in a semantic categorization task are dependent on factors such as how much time the bilinguals use their L2 in day-to-day life, or their self-reported spoken L2 proficiency.

To account for the present patterns, the Sense Model (Finkbeiner et al., 2004) would need to make an additional assumption that knowledge of senses in both languages is being driven by experience as well as semantic representations having a resting-level activation. As learners gain more experience using their L2 in different social interactions and acquire more knowledge about the meanings and uses of words in their L2, not only does one gain more senses that are associated with L2 words, but also that the senses that one has already acquired gradually become more ingrained in memory the more one encounters and uses such senses in conversations. Thus, in tasks such as the semantic categorization task, it would not be sufficient for L2 primes to possess the sense that denotes category membership required to preactivate the category membership of the target. If the resting-level activity of that sense is still low, the activation of L2 representations are still temporally delayed, and the prime cannot preactivate the target. Only once the resting-level activation of the relevant sense becomes higher through active use of the language in the real world can it successfully preactivate the relevant target representations.

Based on these assumptions, the Sense Model (Finkbeiner et al., 2004) could explain Experiment 2 and 3's findings. However, such an account would still have problems with not only the findings of Experiment 1, but also with the results of other studies that have shown effects due to subject proficiency in a lexical decision task (e.g., Nakayama et al., 2016). In a lexical decision task, the core assumption of the Sense Model is that priming is dependent on the ratio of primed-to-unprimed senses. Primes that preactivate a large proportion of the senses associated with the target are predicted to produce significant priming effects, while primes that preactivate only a small proportion of the senses associated with the target are predicted to produce null effects. However, as one gains more experience and knowledge about words in their L2, many of the senses that one would acquire would be language-specific, and should have no effect on L2-L1 priming effects. Given such assumptions, even if the Sense Model were to make the modifications suggested above to account for Experiments 2 and 3's findings, the results of Experiment 1 and of Nakayama et al.'s study are still difficult to reconcile with the Sense Model.

The findings of Experiments 2 and 3 also present an interesting challenge for models such as the BIA+ (Dijkstra & van Heuven, 2002), specifically in how the BIA+ model would account for the dissociation in skills and behaviours associated with L2-L1 translation priming in lexical decision and semantic categorization, particularly the effects of prime frequency on priming effects in each task. The BIA+ model can account for the effects of prime frequency in lexical decision by again assuming that the frequency of the prime affects the general resting-level activation of the prime's representations. Since high-frequency primes have higher resting-level activity, there is less of a delay in accessing the semantics associated with the prime than for low-frequency words. However, this account would have difficulty explaining why prime frequency had a negative relationship with priming in the semantic categorization task, or, for that matter, why the quality of orthographic representations played only a small role in accessing semantics compared to the extent to which learners use their L2 is used in daily life. If access to the prime were simply affected by the resting-level activity of L2 representations, then prime frequency should still have a facilitative effect in semantic categorization. These results show that these effects are constrained by the task context. Further, the factors that predict priming in semantic categorization had little to no positive impact on priming in lexical decision. It remains unclear how a model without a mechanism to allow the task/decision subsystem to exert a top-down influence on processing within the word identification subsystem can demonstrate the computational flexibility required to account for these results.

7.3 The Burden of Specificity Hypothesis

Beyond any of the specific models discussed in relation to the semantic categorization task, I would like to propose the following account of the findings of Experiments 1-3. This account, referred to as the Burden of Specificity Hypothesis (or BSH), argues that the differences observed between the semantic categorization task and the lexical decision task are due to the degree of crispness of lexical representations required for primes to sufficiently activate the relevant representations for targets. Where semantic categorization and lexical decision differ is in the amount of specification of words within the vocabulary required to preactivate the target, specifically with respect to how coherently the meaning-level knowledge of a word is bound to or mapped onto the word's form-level representations (i.e., the word's orthographic and phonological forms; e.g., Perfetti, 2007).

In short, this account assumes that the priming in the L2-L1 direction is contingent on the degree of lexical entrenchment of the L2 required to produce a priming effect, and that the entrenchment required varies from task to task. To produce priming in a lexical decision task, three conditions need to be met. First, L2 learners must have a broad knowledge of the language, which can be operationally defined as the breadth of receptive and productive vocabulary that they have in the language. Second, L2 learners must have well-specified form and meaning representations for the L2 words that are being used in the experiment. Finally, and most critically, the form and meaning components of representations must be well bound together, which is assumed to facilitate the efficient retrieval of the prime's representation.

The binding of form and meaning is assumed to depend on several factors, including the frequency of learners' use of and exposure to their L2 across both the visual and auditory modalities, and word-specific factors such as spoken and written word frequency. When the form- and meaning-level representations are bound only loosely together, the retrieval of the prime's meaning is less efficient, less consistent, and takes a longer period of time. The lexical decision task is assumed to place a premium on how specified the bindings or mappings between form and meaning are for L2 words, specifically with respect to the binding of orthographic and semantic representations. In part, such an explanation is consistent with the finding that writing ability was an important predictor in lexical decision, as writing ability is assumed to reflect several components, including the productive vocabulary of the subject, and orthographic form knowledge in L2. Such an explanation can also account for the effects of prime frequency in the lexical decision task, as the word representations of high-frequency L2 primes would have stronger, more coherent bindings between form- and meaning-level knowledge than low-frequency primes (see Blais, O'Malley, & Besner, 2011, for a theoretical overview of the locus of word frequency effects in word recognition). The spoken comprehension of an L2 would be assumed to have a meaning-level component, as it involves the interpretation of the meaning of information both at the individual word level and at the discourse-level.

In a semantic categorization task, some of the requirements to produce a priming effect overlap with those in the lexical decision task. It is assumed that L2 learners still require a broad knowledge of their L2. Real-world immersion in an L2 environment offers L2 learners an opportunity to gradually accrue more knowledge of their L2 in a naturalistic setting, which helps to broaden learners' grasp of L2 vocabulary and helps learners acquire greater knowledge about the meaning and pragmatic usage of words in their L2. Where the semantic categorization task and the lexical decision task differ is in how specified the mappings between form and meaning need to be to sufficiently preactivate the target. In the semantic categorization task, it has often been suggested that the mechanism that drives translation priming revolves around whether the prime can preactivate conceptual features associated with the target that denotes category membership (e.g., Xia & Andrews, 2015). While the form-meaning mappings would still require some specification to produce priming effects, the requirement is not as high as in lexical decision task, so priming effects can emerge with less-specified mappings than in lexical decision, so long as the meaning-level information that has been bound on to form-level information sufficiently implies the category membership of the target.

The effects of L2 usage in real-world settings in semantic categorization can also be framed in terms of the L2 cultural immersion of the learner, with more frequent use of the L2 in social interactions in an L2 dominant cultural environment reflecting a greater immersion in the L2-dominant culture. Research has suggested that cultural immersion has significant effects on the conceptual representations of bilinguals above and beyond L2 proficiency. In an early study of the effects of cultural immersion, Malt and Sloman (2003) had English L2 learners provide typicality ratings for objects using English. Subjects that spent more time immersed in an L2 cultural environment had typicality ratings that more similar to those of native English speakers, and cultural immersion was a better predictor of native-like ratings than formal instruction.

Critically, the effects of L2 cultural immersion on conceptual representations are not limited to the development of L2 representations. Immersion in an L2 culture can also result in "semantic accents" in their L1, in that the way learners comprehend concepts in their L1 can be influenced by learners' knowledge of the L2 translation equivalent. In a recent study, for example, Matsuki (2018) examined the differential effects of L2 proficiency and L2 cultural immersion on semantic accents in Japanese-English bilinguals' L1 and L2. Matsuki found that bilinguals that had spent more years living in an L2-dominant country had reduced L2 semantic accents, and increased L1 semantic accents. In short, the influence of their knowledge of the L1 translation equivalent on their comprehension of L2 words diminished, while the influence of their knowledge of the L2 translation equivalent on the comprehension of L1 words increased over time.

Although the frequency of L2 usage in social interaction, a multifaceted factor that has been extensively investigated in the present experiments is, of course, not the same as the amount of time spent living in an L2-dominant country. Nonetheless, in certain situations the latter factor may be a good proxy for the former factors in thinking about why the prolonged use of the L2 may affect not only the development of L2 conceptual representations, but also the semantic accenting in L1 representations. The further argument, however, is that it is the use of the L2 that is critical rather than the amount of time that a learner has spent living in an L2-speaking country. Further, using the amount of time that a learner has lived in an L2-speaking country as a measure of cultural immersion may have a major problem, in that it does not account for the possibility that L2 learners may have access to a sizeable community of people who speak their L1. Hence, even though they are living in their L2 country, they may not be exposed to L2 to an extensive degree. For example, the size of the Japanese-speaking community living in Canada is substantially smaller than the size of the Chinese-speaking community (Statistics Canada, 2011). With limited access to an intracultural group to socialize with, Japanese L1 speakers would have fewer opportunities to use their L1, and would spend more of their day-to-day living in an L2-dominant environment. The Chinese-speaking community, however, is sizeable enough that many of their daily social interactions can be done in their L1. As such, the amount of time living in an L2-speaking country may often not be a good approximation of L2 learners' cultural immersion. A better approximation would be obtained from measures of the amount of time that L2 learners actively use their L2[37].

Overall, this account is proposed to provide an explanation for the pattern of results seen in both the lexical decision task (e.g., Gollan et al., 1997; Nakayama et al., 2016), and the semantic

---

[37] As evidence for this idea, when conducting follow-up analyses to examine the effects of the number of years subjects had been living in Canada, the amount of time spent in Canada by subjects weakened the predictions made by sPIP when it was included as a factor in all four experiments.

categorization task (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), through its assumption that the differences in task context place different requirements for how specified the mappings between form and meaning need to be to produce masked translation priming. Several findings are consistent with this account.

First, in lexical decision, the null priming effect seen in the L2-L1 direction tends to be more common in bilinguals whose languages have different scripts (see Schoonbaert et al., 2009, for a meta-analysis), with studies in Hebrew (Gollan et al., 1997), Chinese (Chen et al., 2014), and Japanese (Finkbeiner et al., 2004; however, see Nakayama et al., 2016) all using L1s and L2s with different scripts. Sharing a script can affect processing in several ways. First, as Schoonbaert et al. argued, sharing a script would mean that the early stages of processing would be similar for the two languages, while L2 processing cannot gain benefit from L1 processing when the scripts differ. Further, sharing a script would also mean that the L2 learner has already had a lot of experience with the writing system when learning their L1, allowing subjects to use their already-established form-level knowledge in their L1 as a basis for acquiring lexical orthographic knowledge of their L2, as well as the form mappings between lexical orthography and meaning faster than if they had to additionally become familiarized with a new script.

Second, much of the research that has been done on masked translation priming in contexts where the two languages use different scripts has been done in environments where the required use of the L2 script in daily life is relatively minimal. Specifically, most of the research has been done in countries where subjects are immersed in an L1-dominant social environment, and where most daily activities can be done without the use of their L2. For example, in Gollan et al.'s (1997) study with English-Hebrew and Hebrew-English bilinguals, the Hebrew-English bilinguals were tested in Israel, while the English-Hebrew bilinguals were tested in the United States. Neither of these groups of bilinguals would require the use of their L2 orthography on a consistent basis in daily life. As a result, such subjects would have far less experience with their L2 word forms, and have less opportunity to develop rich mappings between form and meaning in their L2. When bilinguals have been tested in an L2 environment (e.g., Finkbeiner et al., 2004), on the other hand, this research did not consider individual differences in L2 form and meaning knowledge. By averaging over these individual differences instead of accounting for them, such studies may have underestimated L2 learners' abilities to access the meaning of L2 primes in masked translation priming.

Third, such an account can readily explain the lack of a facilitative effect of prime frequency seen in Experiments 2 and 3. If it is assumed that word frequency affects the binding between form- and meaning-level knowledge, and it is further assumed that the semantic categorization task does not require the form-meaning bindings to be as tight to produce priming in the task, as long as the meaning-level information that is bound to form-level knowledge contains information about the category membership of the word, then a robust facilitative effect of prime frequency should not occur in the semantic categorization task.

Fourth, such an account may provide an explanation for why priming effects were larger when the English primes were rated as more typical representations of the target category than when they were rated as more atypical category members, as the more typical English exemplars would be ones that L2 learners would be exposed to the most when living in an L2-dominated environment. For example, L2 learners would be more likely to be exposed to L2 concepts such as *apple*, *orange*, or *banana* than they would *mango*, *fig*, or *coconut*. More typical exemplars would be ones that are more likely to contain sufficient information about the category membership of the target than atypical members.

With respect to the asymmetry observed between L1-L2 and L2-L1 translation priming, this account explains the significant priming effect in the L1-L2 direction in lexical decision as being due to the lexical representations of the L1 primes being crisp, well-specified, and having strong bindings between form and meaning, making the retrieval of lexical representations from the prime efficient enough that the prime can preactivate the representations of the target. Because the lexical representations of the L2 targets are more poorly specified, the processing of these targets is less efficient, providing more opportunity for the prime to influence decisions. In the L2-L1 direction, however, the L2 primes are less specified, and retrieval of the lexical representation is less efficient as a result, reducing the likelihood that the prime will preactivate the target representations. In addition, because the retrieval of the L1 lexical representation is highly efficient, there is less opportunity for the prime to influence the decision. In semantic categorization, the strength of the form-meaning bindings is not as important to the task as it is in lexical decision. What is emphasized, instead, is the semantic information that is bound to the prime, and whether this information is sufficient to activate the target. Even the presence of information that can activate the category membership of the target is sufficient to produce priming, which results in significant L2-L1 priming, but the priming effect is still affected by the overlap between the L1 and L2 translation equivalents at the semantic level. The asymmetry arises because the meaning-level information that is bound onto L2 forms is disproportionately influenced by meaning-level information associated with the L1 translation equivalent. In this circumstance, priming in the L1-L2 direction is robust because L1 primes possess rich semantic representations, and much of the semantic information associated with the more sparsely represented L2 is borrowed from the L1 representation. As a result, even though a priming effect can be obtained in the L2-L1 direction because the basic category-specific information is typically contained by the L2 semantic representation, priming effects would still be larger in the L1-L2 direction.

This account also makes several predictions that can be empirically tested. First, this account predicts that the manner in which the language is learned can affect the time course in which language learners develop priming effects. Under circumstances where the acquisition of the language is similar to that of a native speaker – that is, learners become familiarized with the spoken form of the language before acquiring knowledge of the orthographic forms associated with the language – a trajectory of development should occur in which learners acquire priming effects in the semantic categorization task first. As learners gradually accrue greater knowledge about L2 orthographic forms, and as learners develop more enriched bindings or mappings between these forms and meaning, a priming effect should eventually emerge in the lexical decision task. However, testing such a prediction in bilinguals may prove difficult, as it would require having control over how learners acquired their L2, and would additionally require a longitudinal assessment of those learners over the course of L2 acquisition. An alternative approach, while presumably different from masked translation priming, would be to study the effects of individual differences in L1 reading and writing ability on the development of masked semantic priming in semantic categorization and lexical decision, using a cross-sectional design that examines children and adults at different stages of reading and writing development, and at differing levels of reading and writing skill. Such research would provide useful insight into whether priming in semantic categorization and lexical decision follow the proposed trajectory.

7.4 Translation Priming in Episodic Recognition

When combined with the findings of Experiment 1 showing that translation priming in lexical decision is impacted by L2 listening and writing proficiency, and L2 prime frequency, the results of Experiment 4 cast serious doubts on whether the Episodic L2 Hypothesis (Jiang & Forster, 2001) can adequately account for the translation priming asymmetry observed in lexical decision. Perhaps the most serious issue Experiment 4 presents for the Episodic L2 Hypothesis is its findings regarding the effects of age of L2 acquisition and the number of years learning an L2 on translation priming effects. If it is to be assumed that the L2 is represented in episodic memory when learners acquire the language later in life, as the Episodic L2 Hypothesis argues, this account would have serious difficulty accounting for the fact that priming effects were larger for subjects who acquired their L2 at an earlier age, and who had been learning their L2 for a longer period of time.

Overall, these results suggest that greater L2 proficiency is associated with larger L2-L1 translation priming effects in the speeded episodic recognition task, much like in other experiments. It would be difficult to argue, then, that the translation priming effects observed in speeded episodic recognition tasks (e.g., Jiang & Forster, 2001; Witzel & Forster, 2012) are due to the fact that L2 representations exist solely in the episodic memory system, when other experiments have clearly shown that translation priming effects occur under specific circumstances in tasks which are assumed to require lexical representations in L2. An alternative account must thus be proposed to explain these findings.

One possible explanation for these results is that when subjects study words in their L1, an episodic trace is formed from this encounter. The contents of this trace, however, differ for subjects who are less proficient in their L2 compared to subjects who are more proficient in their L2. For subjects that are proficient in their L2, the memory trace created by exposure to the L1 words contains information about both the L1 word and the L2 word, as a result of both words becoming co-activated upon exposure to the L1 stimulus. When presented with studied targets during the testing phase, the L2 prime can thus aid in the retrieval of the memory trace not because the L2 representations exist solely in episodic memory, as Jiang and Forster (2001) argued, but because the coactivation of the L2 that occurred when the L1 targets are encountered in the study phase produced a trace that also contained the L2 representation. For learners who were less proficient in their L2, the likelihood that the L2 will become coactivated upon exposure to L1 targets is much smaller. As a result, the episodic trace is less likely to contain the L2 representation, and the prime is less likely to aid in the retrieval of the L1 memory trace.

This account is, of course, not without issues. First, the account does not explain why Jiang and Forster (2001) produced priming effects in their episodic recognition task, but not in their lexical decision task. Second, this account explains why Jiang and Forster produced a significant L2-L1 priming effect in episodic recognition, but not in the L1-L2 direction, a result that would appear to be consistent with Jiang and Forster's account. Finally, this account does not easily accommodate Witzel and Forster's (2012) second experiment findings, in which they taught subjects words in a new language, and found that these words produced L2-L1 priming in episodic recognition, but not in lexical decision. At present, the only thing that can be done is speculate as to why the results of the present Experiments 1 and 4, and Jiang and Forster's results were different.

With respect to Jiang and Forster's studies, there are a few issues that need to be considered. First, Jiang and Forster did not systematically study the effects of L2 proficiency on translation priming in either their lexical decision task or their speeded episodic recognition task, nor did they account for the potential impact of item-specific factors like the frequency of the prime on L2-L1 priming. All subjects in Jiang and Forster's experiments were Chinese-English graduate students that had a TOEFL score of 550 or higher, which is considered an average score. However, the authors never systematically studied whether proficiency had an effect on priming in the episodic recognition task or the lexical decision task, nor did they assess the effects of specific dimensions of L2 competency on priming effects like was done in the present research, nor did they perform analyses to assess the effect of prime frequency on translation priming. Thus, it can't be known whether the priming effect in the episodic recognition task and lexical decision task varied as a function of L2 proficiency and item-specific factors based on their results. At the very least, by not systematically accounting for these fine-grained differences between subjects and items and opting to instead look only at the mean RTs, the results of the present research suggest that Jiang and Forster, much like prior research (e.g., Gollan et al., 1997; Grainger & Frenck-Mestre, 1998), did not account for meaningful data in concluding that unbalanced bilinguals cannot produce an L2-L1 translation priming effect in lexical decision, and in concluding that the reason why this null effect occurs in lexical decision is because L2 words are not represented in lexical memory.

A second, but highly related issue with Jiang and Forster's (2001) studies was that the stimuli the authors used were far more homogeneous than the stimuli used in Experiments 1 and 4 in the present research. Jiang and Forster used only high-frequency abstract nouns as targets, and abstract English primes in all of their experiments. Their reasoning was that they wanted to avoid confounding effects of variables such as concreteness. However, in Experiments 1 and 4, one of the goals was to assess whether item-specific factors impact the priming effect produced by systematically studying the combined impact that these factors have on priming using statistical modeling. While still ensuring that each condition had similar mean target frequency, prime frequency, and stroke count in the present research, the increased list size meant that there was less intra-list homogeneity, and more natural variation in both prime and target characteristics, which allowed the present research to also assess the contributions of prime and target lexical characteristics to translation priming by accounting for these differences. By composing their lists of a small, highly homogeneous set of stimuli that represent only a narrow scope of the natural variation that occurs within a language's lexicon, the conclusions that Jiang and Forster drew were likely too broad, given the nature of their stimuli.

A third issue with Jiang and Forster's (2001) studies relates to the number of items used in those studies. Many previous studies that have reported a null L2-L1 priming effect with Chinese-English bilinguals have used underpowered designs, sometimes with fewer than 16 items per cell (e.g., Gollan et al., 1997; Witzel & Forster, 2012; Chen et al., 2014). Jiang and Forster's study was no different from other studies that have reported a null L2-L1 priming effect, as their experiments only used 16 items per cell. At least in the circumstance of lexical decision, a recent meta-analysis by Wen and van Heuven (2017) has shown that the effect size of the L2-L1 translation priming effect is modulated by the number of items per cell. Wen and van Heuven found that studies using a larger number of items per cell produce a larger priming effect than studies using a smaller number of items per cell, a point which was also raised in a recent study by Lee, Jang, and Choi (2018). Further, Brysbaert and Stevens (2018) have recommended that a minimum of 1600 observations per condition is required to achieve the necessary statistical power for these experiments.

Jiang and Forster (2001) would have only had a maximum of 416 observations per condition in their Experiment 1, 256 observations per condition in their Experiment 2, 576 observations per condition in their Experiment 3, 352 observations per condition in their Experiment 4, and only 288 observations per condition in their Experiment 5, before accounting for (and eliminating) error trials. In contrast, Experiments 1 and 2 in the present research had 50 items per cell, and had over 3600 observations per condition. Experiment 3 again had 50 items per cell, and over 1000 observations per condition. Experiment 4 had 60 items per cell in each list, and 120 items per cell when factoring in that lists 1-4 and 5-8 used different sets of stimuli, resulting in around 1800 observations per condition. Combined with the prior issues discussed above, it is likely that Jiang and Forster's experiments were also too underpowered to detect any meaningful differences.

A fourth issue for Jiang and Forster (2001) was that they had subjects perform the episodic recognition task twice because of error rates on the first session, and only analyzed the results of the data from the second session. While such an approach would certainly resolve the issue of high error rates, the issue with such an approach is that it may introduce practice effects that could impact the behavioural results. Experiment 4 did have high error rates for Old trials, but the errors also varied as a function of L2 proficiency, with subjects that reported higher levels of L2 proficiency producing significantly fewer errors than subjects that reported lower levels of L2 proficiency. Thus, a decision was made to not have subjects perform the task twice, because not only would it have required a significantly longer session to complete given the much larger sample of stimuli that were used in Experiment 4, it would have also introduced practice effects.

Finally, one key difference between the present research and Jiang and Forster's (2001) and Witzel and Forster's (2012) research was the concreteness of the stimuli that were used. While the words used in the present research were more heterogeneous across factors such as prime and target frequency than the stimuli used by Forster and colleagues, my items were homogeneous across other factors. One such factor was concreteness. Contrary to prior research by Forster and colleagues, which used strictly abstract words, the present research used mostly concrete words in both episodic recognition and lexical decision. The present research would appear to be the first to use concrete concepts in masked translation priming in episodic recognition, and this distinction may be critical for understanding the difference between the results of the present study and Forster and colleagues' results, as the processing of such stimuli, and indeed the representation of such stimuli within memory, is presumably different. Perhaps the most important distinction between the stimuli in the present research and Forster and colleagues' stimuli is that the stimuli in the present studies would have sensorimotor referents. Paivio's (1971, 1986) Dual Coding Theory (DCT), in particular, argues that concepts can be represented across two modality-specific systems: a nonverbal system that represents the perceptual and sensorimotor characteristics of concepts, and a verbal system that represents concepts using arbitrary linguistic symbols. According to DCT, where concrete and abstract concepts differ is in the modality-specific systems that can be employed when processing and comprehending such concepts. Concrete concepts are assumed to have representations in both the verbal and nonverbal system, and it is further assumed that these verbal and nonverbal representations are mutually interconnected. Abstract concepts, on the other hand, have no nonverbal referent, and processing of such concepts is thus less efficient.

One question that the present research raises, for unbalanced bilinguals at least, is whether the ability to integrate concepts into lexical memory is affected by the types of referents that the concept possesses. For concrete words, such concepts have a variety of visual, auditory, tactile, olfactory, gustatory, and action-based referents associated with them. The concept *apple*, for example, is associated with a large array of sensorimotor information about the concept, including the sight, smell, feel, taste, and any motor-based actions (e.g., grasping, biting) that are associated with the concept. For concepts such as *dignity*, however, no such sensorimotor referents exist. Perhaps, then, having these referents aids in the development of a stable lexical representation? In short, for someone who acquires an L2 at a later stage in life, integrating concepts in an L2 is aided by having tangible referents outside of the arbitrary labels used to denote the concept, making such concepts more likely to eventually transition from episodic to lexical memory. Such an explanation could account for the null overall effect of priming in Experiment 4, and why priming effects were considerably smaller than what was observed by Jiang and Forster (2001) and Witzel and Forster (2012).

One issue with this explanation is that it would still not explain why the largest facilitative factors in the sPIP score derived from Experiment 4 were global L2 proficiency, and the number of years subjects had been acquiring the language. I have gone into some detail about several possible accounts that could explain the results of Experiment 4, but as it stands, there is no account which is unequivocally favoured by the data over the others. A possible solution to this issue is provided below.

7.5 Limitations and Future Directions

There are a few methodological limitations of the present research that are worth noting. First, while using mostly nouns as experimental stimuli, there were some stimuli that were used in Experiments 1 and 4 which were also classified as verbs or adjectives, whereas other studies have used strictly nouns (e.g., Jiang & Forster, 2001). It is possible, then, that the grammatical class of the targets had an impact on the behavioural results obtained for these stimuli. However, it is unlikely that this issue would be a serious one, as the vast majority of stimuli used in these experiments were nouns. Regardless, it should at least be acknowledged that there were verbs and adjectives that were included in the lists. A third issue, as discussed in the General Discussion, was that it is unclear what mechanism could plausibly account for both the results of Experiment 4, and simultaneously the results of Jiang and Forster (2001) and Witzel and Forster's (2012) studies. One avenue that can be taken to improving our current understanding of the representation of L2 in memory is by systematically studying the effects of concreteness on L2-L1 priming in episodic recognition. The present research used mostly concrete concepts, whereas prior research that has studied L2-L1 translation priming in episodic recognition has used abstract concepts. Understanding how concrete and abstract concepts are represented in bilingual memory, then, could provide the necessary insight to properly evaluate the Episodic L2 Hypothesis' ability to accommodate findings from recent lexical decision research (e.g., Nakayama et al., 2016).

One final issue with the present research relates to the use of sPIP, iPIP, and PIP. While these measures were used to compensate for the relative homogeneity in subjects standardized proficiency measures, there are a few issues with these measures. First, the precision and accuracy of these measures were only as good as the factors that they were composed of.

Specifically, there may be factors that were not considered in the present research that significantly impact translation priming. For example, the role of receptive and productive vocabulary size in translation priming is currently not well understood, and was not accounted for in these measurements. Accounting for factors such as individual differences in vocabulary size, then, could improve the precision and generalizability of the sPIP measurement. Future research will need to identify a more comprehensive set of factors which contribute to translation priming to better understand the role that these factors play. For vocabulary size, for example, one approach that should be considered would be using lexical tests such as LexTALE (Lemhofer & Broersma, 2012) to provide estimates of vocabulary size, as such measures have been shown to be good predictors of English vocabulary knowledge, and provide a more accurate measure of English proficiency than self-ratings.

Second, the computation of sPIP largely consisted of self-reported factors. The estimates that were used to make predictions about L2-L1 translation priming effects relied on the accuracy of each subject's self-assessment of their abilities in their L2. Initially, IELTS was intended to be included as a measure in sPIP, but the measure was too homogeneous to reliably distinguish between each subject's actual proficiency in their L2. Access to the individual components of each subject's IELTS score was also limited, rendering the usefulness of the measure limited. Further, due to the limited amount of time in each session, there was not enough time to assess subjects using other objective measures of L2 knowledge. Thus, the initial measure of sPIP was based on subjects' self-reported L2 proficiency. However, future research can improve on this methodology by using more objective measures of L2 proficiency and vocabulary knowledge. One avenue that has already been suggested is in using lexical tests such as LexTALE (Lemhofer & Broersma, 2012), while other avenues may include using tests such as the Nelson-Denny Reading Test (Brown, Fishco, & Hanna, 1993), or using the individual components of scores such as IELTS, TOEIC, or TOEFL as predictive factors, rather than overall scores. Such approaches would provide the advantage of providing a fine-grained approach to understanding the nuanced nature of how L2 proficiency contributes to L2-L1 priming, while retaining the use of objective, standardized measures of L2 proficiency.

Third, the PIP measures are not standardized. Subjects who score on the high end of the sPIP score, for example, are scoring higher on the sPIP score in relation to other subjects in these experiments. It is unknown, however, whether these subjects would score higher on this measure compared to the larger ESL population. The same issue also applies to the iPIP score, and the PIP score as a whole. The scores of these subjects and items can only be evaluated relative to the other subjects and items within the sample. Further, it is also unknown whether the factors derived from Chinese-English bilingual studies would generalize to research using different scripts, languages, and orthographies, such as Hebrew, Korean, or Japanese. It is possible that some of the factors that affected translation priming in lexical decision are specific to the language comparison being used. One goal of future research should be to standardize these measures in a larger scale norming study, using a larger sample of subjects and items, a more comprehensive list of subject- and item-specific factors, and afterwards, a wider variety of language and task comparisons. Such an undertaking was too large in scope to be addressed in the present research. The use of sPIP, iPIP, and PIP in the present research thus represents only the first step towards developing a more sophisticated understanding of the factors that contribute to translation priming, and how these factors differ across different tasks, and potentially, across different language comparisons.

Overall, the present research represents one of the first steps towards accounting for learner- and item-level differences in bilingual language processing. Such an individual differences approach has both its strengths and weaknesses. This approach has provided a useful approach in identifying concise sets of factors that predict behavioural outcomes in experimental tasks, and can be used to demonstrate how these factors differentially affect performance across different experimental tasks, even when the solution to the problem is poorly defined, and the number of potential predictors is large. This approach has also gone beyond looking at global L2 proficiency and has provided a nuanced method of assessing the role of different facets of L2 proficiency in driving translation priming. Such an approach has also been shown to have results that can replicate across different samples, demonstrating the reliability of these factors in predicting behavioural outcomes.

The approach that has been outlined in the present research is, as mentioned, just one step towards developing a more sophisticated method of predicting behavioural outcomes such as translation priming using subject- and item-specific predictors. In continuing to develop this approach, several challenges need to be addressed. First, future research will need to collect a larger and broader sample of predictive measures, such as vocabulary size, to assess how individual differences across these measures contribute to translation priming. The predictive ability of measures such as PIP is only as good as the measures that it is composed of. Second, if this approach is to have any utility in future research, it is necessary that measures such as sPIP and iPIP are normed on a large, diverse sample of subjects and items across a diverse set of tasks to ensure that the factors derived from this approach reliably predict priming outcomes beyond the sample used to fit the measures. Finally, this approach should be taken using a diverse sample of different language comparisons. There may be factors that contribute to translation priming that are language-specific, but of equal interest is whether there are factors that can generalize across languages in how they contribute to bilingual language processing. Such extensive norming was not feasible in the present research, but future collaborative work may help to develop standardized measures that can be used by other researchers.

7.6 Conclusions

The present experiments were an attempt to address the issue of the apparent task-specific nature of the masked translation priming effect that has been reported in prior studies (e.g., Finkbeiner et al., 2004; Gollan et al., 1997; Grainger & Frenck-Mestre, 1998; Jiang & Forster, 2001). Using a machine-learning approach to understand the subject- and item-specific factors which contribute to masked translation priming, the present experiments showed evidence that the factors that contribute to the ability of translation primes to activate the relevant representations of their target are specific to the task that subjects are trying to perform. In lexical decision, priming effects were larger for subjects who reported having better spoken comprehension and writing abilities in English, but weaker reading and writing abilities in Chinese, especially when the Chinese targets were low-frequency, and the English primes were high-frequency. In semantic categorization, priming effects were larger for subjects who reported using English more frequently in daily living, especially when the Chinese targets were high-frequency, and the English primes were low-frequency. In episodic recognition, priming effects were larger for subjects who reported having strong reading, writing, speaking, and listening proficiency in English, and who had been learning English for a longer period of time.

Above all else, the experiments presented in the present dissertation highlight the importance of understanding how individual differences in the proficiencies of L2 learners and item-specific differences contribute to performance in translation priming tasks, and represent a major step towards developing a large-scale, data-driven approach to understanding how bilingual memory processes influence the process of visual word recognition, and how these processes vary according to task demands. Given the results presented, future research should continue to pursue developing more comprehensive data-driven tools to develop a more sophisticated understanding of how second language acquisition affects the development of lexical and conceptual memory for words in both L1 and L2.

References

Abdel Latif, M. M. M. (2013). What do we mean by writing fluency and how can it be validly measured? *Applied Linguistics*, *34*(1), 99-105. DOI: 10.1093/applin/ams073

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *2nd International Symposium on Information Theory* (pp. 267-281). Budapest, Hungary: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. DOI: 10.1109/TAC.1974.1100705

Allyn, F., & Burt, J. S. (1998). Pinch my wig or winch my pig: Spelling, spoonerisms, and other things. *Reading and Writing: An Interdisciplinary Journal*, *10*(1), 51-74.

Altarriba, J. (1991). *Is there automatic semantic priming across a bilingual's two languages?* Paper presented at the meeting of the American Psychological Society, Washington, DC.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition, and Neuroscience*, *31*(7), 940-966. DOI: 10.1080/23273798.2016.1171366

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

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

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*(3), 191-208. DOI: 10.1016/j.jslw.2009.05.003

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1563-1571. DOI: 10.1037/a0032186

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340-397.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283-316. DOI: 10.1037/0096-3445.133.2.283

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meaning in word recognition: A review of the literature. In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187-222). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Banks, B., Gowen, E., Munro, K. J. & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *137*(4), 2015-2024. DOI: 10.1121/1.4916265

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, *35*(5), 953-965. DOI: 10.3758/BF03193468

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. DOI: 10.18637/jss.v067.i01

Bialystok, E. (1997). Effects of bilingualism and biliteracy on children's emerging concepts of print. *Developmental Psychology*, *33*(3), 429-440.

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522-538. DOI: 10.1016/j.jneuroling.2007.07.001

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Science*, *16*(4), 240-250. DOI: 10.1016/j.tics.2012.03.001

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525-531. DOI: 10.1017/S1366728909990423

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858-886.

Brown, J. A., Fishco, V. V., & Hanna, G. (1993). Nelson-Denny reading test: Manual for scoring and interpretation, Forms G & H. Rolling Meadows, IL: Riverside Publishing.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, *13*(3), 359-371. DOI: 10.1017/S1366728909990344

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530-548. DOI: 10.1017/S1366728916000353

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. Journal of Cognition, *2*, 1-20. DOI: 10.5334/joc.10

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodal inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

Burnham, K. P., & Anderson, D. R. (2004). Multimodal inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261-304. DOI: 10.1177/0049124104268644

Burt, J. S., & Fury, M. B. (2000). Spelling in adults: The role of reading skills and experience. *Reading and Writing: An Interdisciplinary Journal*, *13*(1-2), 1-30.

Burt, J. S., & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language*, *46*, 518-543. DOI: 10.1006/jmla.2001.2818

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition*, *28*(1), 143-153.

Chen, B., Zhou, H., Gao, Y., & Dunlap, S. (2014). Cross-language translation priming asymmetry with Chinese-English bilinguals: A test of the sense model. *Journal of Psycholinguistic Research*, *43*(3), 225-240. DOI: 10.1007/s10936-013-9249-3

Chen, H-. C., & Ng, M-. L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, *17*(4), 454-462.

Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2013). Task modulation of brain responses in visual word recognition as studied using EEG/MEG and fMRI.

Coltheart, M., Davelaar, E., Jonasson, J. F. & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and Performance VI (pp. 535-555). Hillsdale, N.J.: Erlbaum.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic stats. *Behavior Research Methods*, *37*(1), 65-70.

de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*(1), 90-123. DOI: 10.1016/0749-596X(91)90012-9

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001-1018. DOI: 10.1037/0278-7393.18.5.1001

de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27-51). Amsterdam: John Benjamins.

de Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*(5), 600-629. DOI: 10.1006/jmla.1994.1029

de Wit, B., & Kinoshita, S. (2014a). Relatedness proportion effects in semantic categorization: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1733-1744. DOI: 10.1037/xlm0000004 de Wit, B., & Kinoshita, S. (2014b). An RT distribution analysis of relatedness proportion effects in lexical decision and semantic categorization reveals different mechanisms. *Memory & Cognition*, *43*(1), 99-110. DOI: 10.3758/s13421-014-0446-6

de Wit, B., & Kinoshita, S. (2015). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1062-1075. DOI: 10.1037/xlm0000074

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Springer, Berlin, Heidelberg.

Dich, N. (2011). Individual differences in the size of orthographic effects in spoken word recognition: The role of listeners' orthographic skills. *Applied Psycholinguistics*, *32*, 169-186. DOI: 10.1017/S0142716410000330

Dijkstra, T., Haga, F., Bijsterveld, A., & Sprinkhuizen-Kuyper, I. (2012). Lexical competition in localist and distributed connectionist models of L2 acquisition. In J. Altarriba & L. Isurin (Eds.), *Memory, language, and bilingualism: Theoretical and applied approaches* (pp. 48-73). Cambridge: Cambridge University Press.

Dijkstra, A., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system. From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175-197. DOI: 10.1017/S1366728902003012

Dijkstra, T., Timmermans, M., & Schreifers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, *42*, 445-464. DOI: 10.1006/jmla.1999.2697

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011a). Two words, one meaning: Evidence of automatic co-activation of two translation equivalents. *Frontiers in Psychology*, *2*, 188. DOI: 10.3389/fpsyg.2011.00188

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011b). Masked translation priming effects with low proficient bilinguals. *Memory & Cognition*, *39*, 260-275. DOI: 10.3758/s13421-010-0004-9

Dockrell, J. E., & Connelly, V. (2009). The impact of oral language skills on the production of written text. *British Journal of Educational Psychology*, *1*(1), 45-62. DOI: 10.1348/000709909X421919

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, *57*(2), 98-107. DOI: 10.1027/1618-3169/a000013

Duyck, W., Van Assche, E., Dreighe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663-679. DOI: 10.1037/0278-7393.33.4.663

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: New evidence from Dutch-French bilinguals. *Experimental Psychology*, *56*(3), 173-179. DOI: 10.1027/1618-3169.56.3.173

Duzan, H., & Shariff, N. S. B. M. (2015). Ridge regression for solving the multicollinearity problem: Review of methods and models. *Journal of Applied Sciences*, *15*, 392-404. DOI: 10.3923/jas.2015.392.404

E-Prime (2004). PSTNet [Computer software].

European Commission Special Eurobarometer (2006). Europeans and their languages.

European Commission Special Eurobarometer (2012). Europeans and their languages.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146. DOI: 10.1017/S0048577202001373

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*(1), 1-22. DOI: 10.1016/j.jml.2004.01.004

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680-698. DOI: 10.1037/0278-7393.10.4.680

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116-124. DOI: 10.3758/BF03195503

Frenck, C., & Pynte, J. (1987). Semantic representation and surface forms: A look at across language priming in bilinguals. *Journal of Psycholinguistic Research*, *16*(4), 383-396. DOI: 10.1007/BF01069290

Géron, A. (2017). *Hands-on machine learning with scikit-learn & TensorFlow: Concepts, tools, and techniques to build intelligent systems.* Sevastopol, CA: O'Reilly.

Gillam, R. B., & Johnston, J. R. (1992). Spoken and written language relationships in language learning and normally achieving school-age children. *Journal of Speech and Hearing Research*, *35*(6), 1303-1315.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1122-1139. DOI: 10.1037/0278-7393.23.5.1122

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787-814. DOI: 10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562-576.

Gordon, P. C., Lowder, M., & Hoedemaker, R. (2016). Reading in normally aging adults. In H. Wright (Ed.), *Cognitive-linguistic processes and aging.* (pp. 165-191). Amsterdam: John Benjamins Publishing.

Grainger, J., & Beauvillain, C. (1988) Associative priming in bilinguals: Some limits of interlingual facilitation effects. *Canadian Journal of Psychology*, *42*(3), 261-273. DOI: 10.1037/h0084193

Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes*, *13*(6), 601-623. DOI: 10.1080/016909698386393

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67-81. DOI: 10.1017/S1366728998000133

Guess, D. (1969). A functional analysis of receptive language and productive speech: Acquisition of the plural morpheme. *Journal of Applied Behavior Analysis*, *2*(1), 55-64. DOI: 10.1901/jaba.1969.2-55

Hakuta, K. (1986). Mirror of language: The debate on bilingualism: New York: Basic Books.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1331-1356. DOI: 10.1037/0096-1523.22.6.1331

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, *58*, 54-59.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67. DOI: 10.2307/1267351

Holm, A., & Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition*, *59*(2), 119-147.

Huttenlocher, J. (1974). The origins of language comprehension. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium*. Oxford, England: Lawrence Erlbaum.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277-288. DOI: 10.1016/j.actpsy.2007.06.003

Jacquet, M., & French, R. M. (2002). The BIA++: Extending the BIA+ to a dynamical distributed connectionist framework. *Bilingualism*, *5*(3), 202-205.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition*, *2*(1), 59-75.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*(1), 32-51. DOI: 10.1006/jmla.2000.2737

Jin, Y-. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills*, *70*(3, Pt 2), 1139-1154. DOI: 10.2466/PMS.70.4.1139-1154

Johnson, M. D., Acevedo, A., & Mercado, L. (2016). Vocabulary knowledge and vocabulary use in second language writing. *TESOL Journal*, *7*(3), 700-715. DOI: 10.1002/tesj.238

Kaushanskaya, M., Yoo, J., & Marian, V. (2011). The effect of second-language experience on native-language processing. *Vigo International Journal of Applied Linguistics*, *8*(1), 55-77.

Keatley, C. W., Spinks, J. A., & de Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory & Cognition*, *22*(1), 70-84.

Keuleers, E., Stevens, M., Mandera, P., & Brysbart, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, *68*(8), 1665-1692.

Kirsner, K., Brown, H. L., Abrol, S., Chadha, N. K., & Sharma, N. K. (1980). Bilingualism and lexical representation. *Quarterly Journal of Experimental Psychology*, *32*(4), 585-594. DOI: 10.1080/14640748008401847

Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 519-539. DOI: 10.1016/S0022-5371(84)90366-0

Kroll, B. M. (1981). Developmental relationship between speaking and writing. In B. M. Kroll & R. J. Vaan (Eds.), *Exploring speaking-writing relationships: Connections and contrasts* (pp. 32-54). Urbana, IL: National Council of Teachers of English.

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169-199). Mahwah, NJ: Lawrence Erlbaum Publishers.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149-174. DOI: 10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 49-71). Malden, MA: Blackwell Publishers.

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373-381. DOI: 10.1017/s176672891000009x

Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307-322. DOI: 10.1093/applin/16.3.307

Lee, Y., Jang, E., & Choi, W. (2018). L2-L1 translation priming effects in a lexical decision task: Evidence from low proficient Korean-English bilinguals. *Frontiers in Psychology*, *9*, 267. DOI: 10.3389/fpsyg.2018.00267

Lemhofer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods*, *44*(2), 325-343. DOI: 10.3758/s13428-011-0146-0

Liang, Z., Hamada, Y., Oba, S., & Ishii, S. (2018). Characterization of electroencephalography signals for estimating saliency features in videos. *Neural Networks*, *105*, 52-64. DOI: 10.1016/j.neunet.2018.04.013

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381-390. DOI: 10.1037/a0014875

Liow, S. J. R. & Poon, K. K. L. (1998). Phonological awareness in multilingual Chinese children. *Applied Psycholinguistics*, *19*(3), 339-362. DOI: 10.1017/S0142716400010213

Lo, S. E., & Andrews, S. E. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reacting time data. *Frontiers in Psychology*, *6*, 1171. DOI: 10.3389/fpsyg.2015.01171

Lowder, M. W., & Gordon, P. C. (2017). Print exposure modulates the effects of repetition priming during sentence reading. *Psychonomic Bulletin & Review*, *24*(6), 1935-1942. DOI: 10.3758/s13423-017-1248-1

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, *2*(1), 49-55.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, *8*, 1164. DOI: 10.3389/fpsyg.2017.01164

Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, *6*, 47-67.

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research*, *37*(3), 141-170. DOI: 10.1007/s10936-007-9064-9

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940-967. DOI: 10.1044/1092-4388(2007/067)

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97-115. DOI: 10.1017/S1366728903001068

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

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102. DOI: 10.1016/0010-0277(87)90005-9

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension, and reading rate. *Journal of Research in Reading*, *31*(3), 273-284. DOI: 10.1111/j.1467-9817.2008.00371.x

Matsuki, E. (2018). Conceptual representation in bilinguals: A feature-based approach. Unpublished dissertation.

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language*, *25*, 431-444.

Meyer, D. E., & Ruddy, M. G. (1974). *Bilingual word recognition: Organization and retrieval of alternative lexical codes*. Paper presented to the Eastern Psychological Association, Philadelphia.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*(2), 267-296. DOI: 10.1037/a0021890

Monaghan, P., Chang, Y. -N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, *93*, 1-21. DOI: 10.1016/j.jml.2016.08.003

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: Item Response Theory analysis of the Author Recognition Test. *Behavior Research Methods*, *47*(4), 1095-1109. DOI: 10.3758/s13428-014-0534-3

Muhammad, I., Maria, J., & Muhammad, A. R. (2013). Comparison of shrinkable regression for remedy of multicollinearity problem. *Middle East Journal of Scientific Research*, *14*(4), 570-579. DOI: 10.5829/idosi.mejsr.2013.14.4.488

Nakayama, M., Ida, K., & Lupker, S. J. (2016). Cross-script L2-L1 noncognate translation priming in lexical decision depends on L2 proficiency: Evidence from Japanese-English bilinguals. *Bilingualism: Language and Cognition*, *19*(5), 1001-1022. DOI: 10.1017/S1366728915000462

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese-English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes*, *27*, 1563-1583. DOI: 10.1080/01690965.2011.606669

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2013). Masked translation priming with Japanese-English bilinguals: Interactions between cognate status, target frequency, and L2 proficiency. *Journal of Cognitive Psychology*, *25*, 949-981. DOI: 10.1080/20445911.2013.839560

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1-18). New York: Plenum Press.

Ordonez, C. L., Carlo, M. S., Snow, C. E., & McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology*, *94*(4), 719-728. DOI: 10.1037/0022-0663.94.4.719

Oyeyemi, G. M., Ogunjobi, E. O., & Folorunsho, A. I. (2015). *International Journal of Statistics and Applications*, *5*(2), 72-76. DOI: 10.5923/j.statistics.20150502.04

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

Paivio, A. (1971). *Imagery and verbal processes*. Hillsdale, NJ: Erlbaum.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York, NY: Oxford University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,…Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Pennington, B. F., Lefly, D. L., Van Orden, G. C., Bookman, M. O., & Smith, S. D. (1987). Is phonology bypassed in normal or dyslexic development? *Annals of Dyslexia*, *37*(1), 62-89. DOI: 10.1007/BF02648060

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357-383. DOI: 10.1080/10888430701530730

Perfetti, C., & Adlof, S. M. (2012). Reading comprehension: A conceptual framework from word meaning to text meaning. In J. P. Sabatini, A. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability*. (pp. 3-20). Lanham, MD: Rowman & Littlefield Education.

Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeve, C. Elbron, & P. Reitsma (Eds.), *Precursors of functional literacy*. (pp. 189-213). Amsterdam: John Benjamins.

Pexman, P. M., & Lupker, S. J. (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, *53*(4), 323-334. DOI: 10.1037/h0087320

Python Software Foundation (2015). Python language reference, version 3.5.1. Available at http://www.python.org

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria. http://www.R-project.org/

Randsell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage. *Journal of Memory and Language*, *26*(4), 392-405. DOI: 10.1016/0749-596X(87)90098-2

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245-266. DOI: 10.1006/jmla.2001.2810

Rodriguez-Aranda, C., & Jakobsen, M. (2011). Differential contribution of cognitive and psychomotor functions to the age-related slowing of speech production. *Journal of the International Neuropsychological Society*, *17*(5), 807-821. DOI: 10.1017/S1355617711000828

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*(3), 192-233.

Ruts, W., De Deyne, S., Ameel, E., Vanpaempel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 506-515.

Sanchez-Casas, R. M., Garcia-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing; Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, *4*(4), 293-310. DOI: 10.1080/09541449208406189

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, *9*(3), 259-291. DOI: 10.1016/S1060-3743(00)00028-X

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language and a second language and back: Making sense of the findings. *Memory & Cognition*, *37*(5), 569-586. DOI: 10.3758/MC.37.5.569

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language*, *25*(5), 605-618. DOI: 10.1016/0749-596X(86)90014-8

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*, 197-212. DOI: 10.1016/j.jml.2006.03.004

Snellings, P., van Gelderen, A., & de Glopper, A. (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing*, *21*(2), 174-201. DOI: 10.1191/0265532204lt276oa

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

Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139-152.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, *20*(1), 51-68.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402-433.

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811-826. DOI: 10.1037/0012-1649.31.5.811

Statistics Canada (2011). Census of population, 2011.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.

Taylor, J. N., & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing*, *29*(6), 1069-1103. DOI: 10.1007/s11145-015-9616-6

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267-288.

Tikhonov, A. N. (1943). On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, *39*(5), 195-198.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics*, *4*, 1035-1038.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solution of ill-posed problems*. Washington, DC: Winston & Sons.

Tse, C. S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods*, *49*(4), 1503-1519. DOI: 10.3758/s13428-016-0810-5

Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, *12*(6), 587-595.

van Hell, J. G. (2002). Bilingual word recognition beyond orthography: On meaning, linguistic context, and individual differences. *Bilingualism: Language and Cognition*, *5*(3), 209-212. DOI: 10.1017/S1366728902243011

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431-451. DOI: 10.1016/j.actpsy.2008.03.010

van Heuven, W. J. B., & Dijkstra, T. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458-483. DOI: 10.1006/jmla.1998.2584

Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, *19*(4), 452-470. DOI: 10.1093/applin/19.4.452

Wagenmakers, E. -J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779-804.

Wang, M., Perfetti, C. A., & Liu, Y. (2005). Chinese-English biliteracy acquisition: Cross-language and writing system transfer. *Cognition*, *97*(1), 67-88. DOI: 10.1016/j.cognition.2004.10.001

Wang, X., & Forster, K. I. (2010). Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism and Cognition*, *13*(3), 327-340. DOI: 10.1017/S1366728909990502

Wen, Y., & van Heuven, W. J. B. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review*, *24*, 879-886. DOI: 10.3758/s13423-016-1151-1

West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*(5), 325-330. DOI: 10.1111/j.1467-9280.1991.tb00160.x

Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The Episodic L2 Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1608-1621. DOI:10.1037/a0028072

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). Second language development in writing: Measures of fluency, accuracy and complexity. University of Hawaii Press.

Xia, V., & Andrews, S. (2015). Masked translation priming asymmetry in Chinese-English bilinguals: Making sense of the Sense Model. *The Quarterly Journal of Experimental Psychology*, *68*(2), 294-325. DOI: 10.1080/17470218.2014.944195

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2012). Lexical and sub-lexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1069-1075. DOI: 10.1037/a0026935

Yap, M. J., Balota, D. A., Tse, C. -S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 495-513. DOI: 10.1037/0278-7393.34.3.495

Yap, M. J., Tse, C. -S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency: The role of lexical integrity. *Journal of Memory and Language*, *61*(3), 303. DOI: 10.1016/j.jml.2009.07.001

Zhong, H. F. (2016). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*. DOI: 10.1080/09571736.2015.1127403

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, *67*, 301-320.

Appendix A

Cost Function and Hyperparameter Descriptions for Machine Learning Models

Ridge Regression

Cost Function. For the cost function, the formula takes this form:

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$

If w represents the vector of feature weights θ_1 to θ_n, the regularization term is equal to $\frac{1}{2}(\| w \|_2)^2$, where $\| \cdot \|_2$ represents the sum of squares of the coefficients associated with each vector, also known as the ℓ_2 norm of the weight vector. Finally, the closed form solution is represented as $\hat{\theta} = (X^T \cdot X + \alpha A)^{-1} \cdot X^T \cdot y$, where A is the n x n identity matrix.

Hyperparameters. The first hyperparameter (see Appendix A for a definition of hyperparmeters), α, represents the regularization strength. Larger values of α mean that the coefficients of the predictors in the model will tend to be smaller. When α = 0, the cost function of the model is identical to the cost function of a linear regression without any regularization.

The second hyperparameter, fit_intercept, is a Boolean hyperparameter that is set to True or False. When set to True, the model calculates the intercept. When set to False, the model does not calculate the intercept. The intercept only needs to be calculated when the dependent variable is not centred.

The third hyperparameter, tol, or the convergence tolerance, reflects the required precision of the solution, and is represented as a floating point value. Convergence is defined as the process of arriving at a solution that is as close to the exact solution as possible, using an error tolerance that is pre-specified. The convergence tolerance is best understood using an example. Assume that there is a function $f(x)$ that we need to determine the minimum of. To determine the minimum of $f(x)$, the starting point of the function has already been determined to be x_0, and the way of calculating the gradient $\nabla f(x)$ is already known. To define a successful convergence, we can argue that the algorithm has converged when $|f(x_t) - f(x_{t-1})| < \varepsilon$, where $f(x_t)$ represents the cost at iteration t, $f(x_{t-1})$ represents the cost at iteration $t-1$, and ε represents the convergence tolerance, and is a value greater than zero.

The fourth hyperparameter, copy_X, is an optional Boolean hyperparameter, which, when set to True, copies the values of X. Otherwise, the values of X can be overwritten. The fifth hyperparameter, random_state, is an optional hyperparameter which can either set as an integer value, RandomState instance, or None. This hyperparameter is a seed of a pseudo-random number generator, which is responsible for selecting a random feature to update. If random_state takes on an integer value, random_state is the seed used by the random number generator. If random_state is set as a RandomState instance, it is treated as the random number generator. Finally, if random_state is set as None, the random number generator is the RandomState instance used by NumPy's[38] random function.

The final hyperparameter, solver, reflects the solver used in the computations. Solver is a hyperparameter that is specific to ridge regression. There are seven options for this hyperparameter. First, is 'auto', which chooses the solver automatically based on the type of data. The second is 'svd', which uses a Singular Value Decomposition of X to calculate the coefficients. The third is 'cholesky', which uses a standard scipy.linalg.solve[39] function to find a closed-form solution. The fourth, 'sparse_cg', uses a conjugate gradient solver. The fifth, 'lsqr', uses a regularized least-squares routine. Finally, 'sag' uses a Stochastic Average Gradient

[38] NumPy (numerical Python) is a package for scientific computing in Python, allowing one to create N-dimensional arrays, use linear algebra, and generate random numbers.
[39] SciPy (Scientific Python) is a package for mathematics, science, and engineering. The function mentioned is one of its linear algebra functions.

descent, and ‘saga’ uses an improved version of ‘sag’. The summary of hyperparameters set for the ridge regressions is as follows:

	Experiment									
	Experiment 1		Experiment 2		Experiment 3		Experiments 2 & 3		Experiment 4	
Hyperparameter	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP
α	995.50	211.89	994.00	196.40	12.99	57.97	998.50	240.38	54.47	995.50
copy_X	True	True	True	True	True	True	True	True	True	True
fit_intercept	False	False	False	False	False	False	False	False	False	False
random_state	None	None	None	None	None	None	None	None	None	None
solver	auto	auto	auto	auto	auto	auto	auto	auto	auto	auto
tol	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001

Lasso Regression

Cost Function. The cost function of the lasso regression takes on the following form:

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^{n} |\theta_i|$$

The regularization term for the lasso regression is computed as the sum of the coefficients associated with each vector, multiplied by the α hyperparameter, which is also referred to as the ℓ_1 norm of the weight vector.

Hyperparameters. Many of the hyperparameters that were tuned in the ridge regression were also tuned in the lasso regression, including α, copy_X, fit_intercept, and random_state. In addition, there were also several hyperparameters that the lasso regression and the elastic net regression had that the ridge regression did not. The first of these hyperparameters, precompute, is used to determine whether a precomputed Gram matrix should be used to speed up computations. The lasso regression had this hyperparameter set to True.

The second hyperparameter that was unique to the lasso and elastic net regressions, warm_start, is an optional Boolean hyperparameter, that, when set to True, the model reuses the solution of the previous call to initialize the fitting process. When set to False, calling the model again erases the prior solution.

The third such hyperparameter, positive, is another optional Boolean hyperparameter, that, when set to True, the coefficients of the model are forced to be positive.

The fourth such hyperparameter, selection, selects what coefficients are updated at every iteration, and, when set to ‘random’, causes the model to randomly select a coefficient to update. If selection is set to ‘cyclic’, which is the default, coefficients are looped over sequentially. The hyperparameter values for the lasso regressions are summarized as follows:

	Experiment									
	Experiment 1		Experiment 2		Experiment 3		Experiments 2 & 3		Experiment 4	
Hyperparameter	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP
α	8.17	12.35	6.57	12.35	1.00	5.88	10.16	7.77	1.20	20.31
copy_X	True	True	True	True	True	True	True	True	True	True
fit_intercept	False	False	False	False	False	False	False	False	False	False
positive	False	False	False	False	False	False	False	False	False	False
precompute	True	True	True	True	True	True	True	True	True	True
random_state	None	None	None	None	None	None	None	None	None	None
selection	cyclic	cyclic	cyclic	cyclic	cyclic	cyclic	Cyclic	cyclic	cyclic	cyclic
tol	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
warm_start	True	True	True	True	True	True	True	True	True	True

Elastic Net Regression

Cost Function. The cost function of an elastic net takes the following form:

$$J(\theta) = MSE(\theta) + r\alpha \sum_{i=1}^{n} |\theta_i| + \frac{1-r}{2}\alpha \sum_{i=1}^{n} \theta_i^2$$

The first regularization term in this function represents the ℓ_1 norm of the weight vector, which is shares with the lasso regression, and the second regularization term represents the ℓ_2 norm of the weight vector, which it shares with the ridge regression. For this cost function, the parameter *r* represents a mix ratio, and controls how similar the model is to a ridge regression or a lasso regression. When $r = 0$, the model is identical to a ridge regression, while the model is identical to a lasso regression when $r = 1$.

Hyperparameters. The elastic net regression used the same hyperparameters as the lasso regression, with one exception: the l1_ratio hyperparameter. This hyperparameter represents the *r* parameter in the cost function. The model is identical to a ridge regression when l1_ratio = 0, and is identical to a lasso regression when l1_ratio = 1. When 0 < l1_ratio < 1, the penalty is a combination of ℓ_1 and ℓ_2. The hyperparameter values for the elastic net regression were set as follows:

	Experiment									
	Experiment 1		Experiment 2		Experiment 3		Experiments 2 & 3		Experiment 4	
Hyperparameter	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP	sPIP	iPIP
α	103.40	6.30	8.30	6.50	0.40	1.90	59.40	3.40	0.10	7.20
copy_X	True	True	True	True	True	True	True	True	True	True
fit_intercept	False	False	False	False	False	False	False	False	False	False
positive	False	False	False	False	False	False	False	False	False	False
precompute	True	True	True	True	True	True	True	True	True	True
random_state	None	None	None	None	None	None	None	None	None	None
selection	cyclic	cyclic	cyclic	cyclic	Cyclic	cyclic	cyclic	Cyclic	cyclic	cyclic
tol	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
warm_start	True	True	True	True	True	True	True	True	True	True
l1_ratio	.34	.50	.50	.48	.10	.48	.46	.12	.42	.10

Appendix B

Materials used in Experiment 1

Translation Prime	Control Prime	Target
advice	heron	忠告
border	table	边境
chance	safety	机会
dance	bicycle	舞蹈
government	energy	政府
land	soldier	土地
parrot	pocket	鹦鹉
quail	rope	鹌鹑
secret	campaign	秘密
theory	legend	理论
beach	bottle	海滩
candle	dive	蜡烛
college	lane	学院
energy	government	能源
hotel	problem	旅馆
minute	captain	分钟
poetry	carpet	诗歌
road	college	道路
steam	vulture	蒸汽
vote	window	投票
beard	coffee	胡子
captain	minute	队长
comedy	reward	喜剧
forest	bridge	森林
idea	country	理念
mirror	lunch	镜子
post	handsome	岗位
rope	quail	绳子
sunset	luck	夕阳
vulture	steam	秃鹰
album	traffic	专辑
bottle	sand	瓶子
cliff	game	悬崖
discussion	clown	讨论
guitar	career	吉他
legend	theory	传说

pencil	bacon	铅笔
record	metal	记录
sink	camera	水槽
ticket	piano	车票
author	penguin	作者
bridge	forest	桥梁
clock	beer	时钟
dive	candle	潜水
handsome	post	英俊
luck	sunset	运气
penguin	author	企鹅
reptile	sponge	爬虫
skate	swan	滑冰
toilet	turkey	厕所
beer	clock	啤酒
car	instinct	汽车
computer	season	电脑
friend	wall	朋友
instinct	car	直觉
morning	business	早上
problem	hotel	问题
safety	chance	安全
swan	skate	天鹅
wall	friend	墙壁
bicycle	dance	单车
career	guitar	事业
country	idea	国家
game	cliff	游戏
juice	profit	果汁
music	research	音乐
profit	juice	利润
salt	doctor	食盐
table	border	桌子
whistle	customer	哨子
bacon	pencil	咸肉
business	morning	商业
clown	discussion	丑角
doctor	salt	医生
health	kitchen	健康

lunch	mirror	午餐
piano	ticket	钢琴
research	music	研究
soldier	land	军人
traffic	album	交通
bank	tape	银行
bill	glass	法案
camera	sink	相机
coffee	beard	咖啡
dollar	puppet	美元
heron	advice	白鹭
metal	record	金属
pocket	parrot	口袋
reward	comedy	奖励
sponge	reptile	海绵
turkey	toilet	火鸡
carpet	poetry	地毯
customer	whistle	顾客
glass	bill	玻璃
kitchen	health	厨房
neighborhood	voice	邻里
puppet	dollar	木偶
season	computer	季节
tape	bank	胶带
window	vote	窗口

Appendix C

Materials used in Experiments 2 and 3

Translation Prime	Control Prime	Target
bat	sweater	蝙蝠
camel	centipede	骆驼
cow	watermelon	母牛
fox	fly	狐狸
goat	shoulder	山羊
hedgehog	moth	刺猬
hippopotamus	beetle	河马
kangaroo	lemon	袋鼠
lion	locust	狮子
monkey	throat	猴子
mouse	tooth	老鼠
orangutan	nose	猩猩
panda	eye	熊猫
rabbit	scarf	兔子
rhino	tie	犀牛
seal	coat	海豹
squirrel	wasp	松鼠
tiger	banana	老虎
whale	cicada	鲸鱼
zebra	plum	斑马
ant	chest	蚂蚁
bee	blouse	蜜蜂
beetle	hippopotamus	甲虫
butterfly	cherry	蝴蝶
caterpillar	lips	毛虫
centipede	camel	蜈蚣
cicada	whale	蝉鸣
cockroach	olive	蟑螂
cricket	onion	蟋蟀
dragonfly	belt	蜻蜓
earwig	eyeglasses	蜈蚣
flea	necklace	跳蚤
fly	fox	苍蝇
grasshopper	gloves	蚱蜢
locust	lion	蝗虫
louse	mushroom	头虱

mantis	slippers	螳螂
mosquito	pear	蚊虫
moth	hedgehog	飞蛾
wasp	squirrel	黄蜂
apple	apron	苹果
banana	tiger	香蕉
beet	muscle	甜菜
celery	pyjamas	芹菜
cherry	butterfly	樱桃
corn	chin	玉米
cucumber	crown	黄瓜
grape	hat	葡萄
lemon	kangaroo	柠檬
lettuce	pancreas	生菜
mushroom	louse	冬菇
olive	cockroach	橄榄
onion	cricket	洋葱
orange	stomach	橙子
pear	mosquito	鸭梨
pineapple	liver	菠萝
plum	zebra	李子
strawberry	heart	草莓
tomato	arm	番茄
watermelon	cow	西瓜
apron	apple	围裙
belt	dragonfly	腰带
blouse	bee	衬衫
boots	thumb	靴子
bra	skin	胸罩
coat	seal	上衣
crown	cucumber	皇冠
eyeglasses	earwig	眼镜
gloves	grasshopper	手套
hat	grape	帽子
necklace	flea	项链
pyjamas	celery	睡衣
sandals	back	凉鞋
scarf	rabbit	围巾
shoes	ear	鞋子

skirt	skull	短裙
slippers	mantis	拖鞋
socks	finger	袜子
sweater	bat	毛衣
tie	rhino	领带
arm	tomato	胳膊
back	sandals	背部
chest	ant	胸部
chin	corn	下巴
ear	shoes	耳朵
eye	panda	眼睛
finger	socks	手指
heart	strawberry	心脏
lips	caterpillar	嘴唇
liver	pineapple	肝脏
muscle	beet	肌肉
nose	orangutan	鼻子
shoulder	goat	肩膀
skin	bra	皮肤
skull	skirt	头骨
stomach	orange	肠胃
throat	monkey	喉咙
thumb	boots	拇指
tongue	leopard	舌头
tooth	mouse	牙齿

Appendix D

Materials used in Experiment 4, Lists 1-4

Translation Prime	Control Prime	Target
accident	apartment	意外
apartment	accident	公寓
ball	carpet	球类
carpet	ball	地毯
colour	courage	颜色
courage	colour	勇气
customer	dolphin	顾客
dolphin	customer	海豚
ribbon	footprint	丝带
footprint	ribbon	足迹
green	kettle	绿色
kettle	green	水壶
lecture	maze	讲座
maze	lecture	迷宫
movie	pink	影片
pink	movie	粉红
reptile	stable	爬虫
snail	vote	蜗牛
stable	reptile	马棚
vote	snail	投票
animal	bail	动物
bail	animal	保释
captain	college	队长
college	captain	大学
cook	crystal	厨师
crystal	cook	水晶
dive	emotion	潜水
emotion	dive	情感
flower	gem	花朵
gem	flower	宝石
idea	magazine	理念
law	pepper	法规
magazine	idea	杂志
moral	record	道德
pepper	law	胡椒
record	moral	记录

sign	square	路标
square	sign	方块
variety	workshop	品种
workshop	variety	工场
almond	bacon	杏仁
bacon	almond	咸肉
camera	citizen	相机
citizen	camera	公民
computer	cream	电脑
cream	computer	奶油
dinner	ear	晚餐
ear	dinner	亲耳
floor	garden	地板
garden	floor	花园
honey	knight	蜜糖
knight	honey	骑士
lobster	million	龙虾
million	lobster	万般
past	race	往事
race	past	种族
sheep	spring	绵羊
spring	sheep	春季
truth	witch	真相
witch	truth	女巫
acne	aunt	粉刺
aunt	acne	大妈
beard	ceremony	胡子
ceremony	beard	仪式
comics	court	漫画
court	comics	法院
dentist	doorway	牙医
doorway	dentist	门口
finance	forest	金融
forest	finance	森林
guest	kitchen	客人
kitchen	guest	厨房
letter	metal	信件
metal	letter	金属
nurse	plant	护士

plant	nurse	植物
seed	sponge	种子
sponge	seed	海绵
trip	wire	旅程
wire	trip	电线
actor	baseball	演员
baseball	actor	棒球
bouquet	coconut	花束
coconut	bouquet	椰子
class	fruit	阶级
fruit	class	水果
highway	instinct	公路
instinct	highway	直觉
map	music	地图
music	map	音乐
officer	penguin	官员
penguin	officer	企鹅
picnic	reason	野餐
reason	picnic	理性
route	secret	路线
secret	route	秘密
skill	symbol	技能
symbol	skill	符号
toilet	vest	厕所
vest	toilet	背心
octopus	bottle	章鱼
bottle	octopus	奶瓶
clown	earth	丑角
earth	clown	地球
football	grain	足球
grain	football	粮食
hotel	limit	旅馆
limit	hotel	极限
message	novel	讯息
novel	message	新奇
pearl	piano	珍珠
piano	pearl	钢琴
poetry	rope	诗歌
rope	poetry	绳子

seafood	signal	海鲜
signal	seafood	信号
summer	thigh	夏天
thigh	summer	大腿
union	wisdom	工会
wisdom	union	智慧
autumn	bill	秋季
bill	autumn	法案
chain	desire	连环
desire	chain	愿望
foam	gold	泡沫
gold	foam	黄金
horizon	length	眼界
length	horizon	长度
medal	nails	勋章
nails	medal	指甲
park	photo	公园
photo	park	照片
plateau	road	高原
road	plateau	道路
scene	shadow	画面
shadow	scene	影子
steak	tennis	牛排
tennis	steak	网球
tragedy	whistle	悲剧
whistle	tragedy	哨子
airport	battery	机场
battery	airport	电池
cake	crayon	蛋糕
crayon	cake	蜡笔
fan	furniture	风扇
furniture	fan	家具
homework	laser	功课
laser	homework	激光
mayor	mustard	市长
mustard	mayor	芥末
oxygen	perfume	氧气
perfume	oxygen	香水
pirate	reward	海盗

reward	pirate	奖励
salty	service	咸味
service	salty	侍候
stage	tape	阶段
tape	stage	胶带
traffic	week	交通
week	traffic	星期
album	bread	专辑
bread	album	面包
chest	coffee	胸部
coffee	chest	咖啡
depth	elephant	深度
elephant	depth	大象
glass	import	玻璃
import	glass	进口
license	mango	执照
mango	license	芒果
pain	pillow	疼痛
pillow	pain	枕头
privacy	roadblock	隐私
roadblock	privacy	路障
sisters	spirit	姊妹
spirit	sisters	精神
steam	subway	蒸汽
subway	steam	地铁
vulture	weight	秃鹰
weight	vulture	重量
beach	chance	海滩
chance	beach	机会
cliff	debt	悬崖
debt	cliff	债务
drama	feast	戏剧
feast	drama	盛宴
history	language	历史
language	history	语言
mail	mission	邮件
mission	mail	使命
peanut	priest	花生
priest	peanut	牧师

rent	sink	租金
sink	rent	水槽
soldier	star	军人
star	soldier	星星
strength	voice	力量
voice	strength	声音
wave	yellow	波浪
yellow	wave	黄色
balloon	champagne	气球
champagne	balloon	香槟
circle	curve	圈子
curve	circle	曲线
dollar	excellent	美元
excellent	dollar	优秀
hairstyle	judge	发型
judge	hairstyle	法官
loss	mirror	亏损
mirror	loss	镜子
peak	post	高峰
post	peak	岗位
religion	screw	宗教
screw	religion	螺丝
slush	stairs	烂泥
stairs	slush	楼梯
storm	teacher	风暴
teacher	storm	师傅
waterfall	writing	瀑布
writing	waterfall	笔迹
backpack	button	背包
button	backpack	按钮
chores	cotton	家务
cotton	chores	棉花
doctor	elevator	医生
elevator	doctor	电梯
gossip	interest	八卦
interest	gossip	趣味
lipstick	miracle	唇膏
miracle	lipstick	奇迹
painting	police	绘画

police	painting	警方
rainbow	science	彩虹
science	rainbow	理科
skate	spot	滑冰
spot	skate	斑点
stone	sunrise	石头
sunrise	stone	日出
wallet	worry	钱包
worry	wallet	心事

Appendix E

Materials used in Experiment 4, Lists 5-8

Translation Prime	Control Prime	Target
angel	candy	天使
candy	angel	糖果
crisis	dismissal	危机
dismissal	crisis	解雇
error	food	错误
food	error	食物
garage	handsome	车房
handsome	garage	英俊
holiday	jail	假期
jail	holiday	监狱
list	mood	清单
mood	list	心情
orange	poker	橙色
poker	orange	扑克
rose	shock	玫瑰
shock	rose	电击
stamp	talks	邮票
talks	stamp	会谈
tofu	vapour	豆腐
vapour	tofu	蒸气
breath	comb	气息
comb	breath	梳子
disease	energy	疾病
energy	disease	能源
farm	future	农场
future	farm	未来
ground	helmet	地面
helmet	ground	头盔
husband	king	老公
king	husband	国王
month	oil	月份
oil	month	石油
peace	rhythm	和平
rhythm	peace	节奏
share	smile	股份
smile	share	笑容

swamp	theory	沼泽
theory	swamp	理论
universe	yacht	宇宙
yacht	universe	游船
author	century	作者
century	author	世纪
diamond	drunk	钻石
drunk	diamond	酒鬼
family	frog	家庭
frog	family	青蛙
gray	height	灰色
height	gray	高度
hospital	juice	医院
juice	hospital	果汁
mars	oatmeal	火星
oatmeal	mars	麦片
outcome	price	结局
price	outcome	物价
sewage	size	污水
size	sewage	尺码
summary	theatre	提要
theatre	summary	戏院
tunnel	winner	隧道
winner	tunnel	赢家
art	cartridge	文艺
cartridge	art	墨盒
demon	dose	恶魔
dose	demon	剂量
event	form	事件
form	event	形式
glory	harbour	光荣
harbour	glory	海港
hope	jewelry	希望
jewelry	hope	首饰
luck	myth	运气
myth	luck	神话
order	pool	秩序
pool	order	泳池
schedule	shop	日程

shop	schedule	商店
street	taste	街头
taste	street	口味
tomorrow	wedding	明日
wedding	tomorrow	婚礼
adult	bathroom	成人
bathroom	adult	浴室
career	church	事业
church	career	教会
contest	dream	赛事
dream	contest	梦想
fate	game	命运
game	fate	游戏
hero	lunch	英雄
lunch	hero	午餐
meat	neighbour	肉类
neighbour	meat	邻居
patient	profit	病人
profit	patient	利润
residence	soup	住处
soup	residence	汤水
syrup	ticket	糖浆
ticket	syrup	车票
utensil	window	用具
window	utensil	窗口
baby	calendar	婴儿
calendar	baby	月历
chicken	cold	鸡肉
cold	chicken	寒意
dawn	fat	破晓
fat	dawn	脂肪
filth	heaven	秽物
heaven	filth	天堂
light	mask	光线
mask	light	口罩
nature	paint	性质
paint	nature	油漆
pleasure	refugee	乐趣
refugee	pleasure	难民

song	surface	歌曲
surface	song	表面
tent	turkey	帐蓬
turkey	tent	火鸡
wife	youth	老婆
youth	wife	青春
arena	border	擂台
border	arena	边境
champion	clock	冠军
clock	champion	时钟
dance	engine	舞蹈
engine	dance	引擎
ferry	heartbeat	渡轮
heartbeat	ferry	心跳
ink	market	墨水
market	ink	销路
menu	ocean	菜单
ocean	menu	海洋
pencil	recipe	铅笔
recipe	pencil	食谱
slope	sunset	坡度
sunset	slope	日落
team	truck	团队
truck	team	货车
wheat	yoga	小麦
yoga	wheat	瑜伽
alcohol	cartoon	酒精
beer	climate	啤酒
cartoon	alcohol	卡通
climate	beer	气候
corner	emperor	角落
emperor	corner	皇帝
feature	ham	特点
ham	feature	火腿
idol	magic	偶像
magic	idol	魔法
media	nerve	传媒
nerve	media	神经
pattern	puppet	格局

puppet	pattern	木偶
scholar	sun	学者
sun	scholar	太阳
tattoo	toy	纹身
toy	tattoo	玩具
watch	wonton	手表
wonton	watch	馄饨
advice	assistant	忠告
assistant	advice	帮手
benefit	brothers	效益
brothers	benefit	兄弟
chariot	comedy	战车
comedy	chariot	喜剧
diary	ecology	日记
ecology	diary	生态
friend	gift	朋友
gift	friend	礼物
legend	palace	传说
palace	legend	殿堂
pork	purple	猪肉
purple	pork	紫色
secretary	sofa	书记
sofa	secretary	沙发
squid	surgery	鱿鱼
surgery	squid	手术
tourist	umbrella	游客
umbrella	tourist	雨伞
angle	bar	角度
bar	angle	酒吧
breakfast	casino	早餐
casino	breakfast	赌场
chili	dessert	辣椒
dessert	chili	甜品
drug	fiction	毒品
fiction	drug	小说
galaxy	heron	银河
heron	galaxy	白鹭
minute	plastic	分钟
plastic	minute	塑胶

prisoner	scale	囚徒
scale	prisoner	规模
sketch	space	草图
space	sketch	空间
story	table	故事
table	story	桌子
train	weather	火车
weather	train	天气
ankle	bathtub	脚踝
bathtub	ankle	浴缸
bridge	clothing	桥梁
castle	devil	城堡
clothing	bridge	服装
devil	castle	魔鬼
earring	flaw	耳环
flaw	earring	破绽
gasoline	hypnosis	汽油
hypnosis	gasoline	催眠
mystery	pocket	玄机
pocket	mystery	口袋
pulse	season	脉搏
season	pulse	季节
snowflake	tire	雪花
spark	student	火花
student	spark	学生
tire	snowflake	轮胎
tribe	winter	部族
winter	tribe	冬季
air	attitude	空气
attitude	air	态度
bicycle	cabin	单车
cabin	bicycle	小屋
cheese	cookie	乳酪
cookie	cheese	饼干
drink	empire	饮料
empire	drink	帝国
function	guitar	职能
guitar	function	吉他
midnight	pharmacy	半夜

pharmacy	midnight	药房
poster	red	海报
red	poster	红色
silver	spa	银色
spa	silver	温泉
sticker	swan	贴纸
swan	sticker	天鹅
tradition	virus	传统
virus	tradition	病毒

Appendix F

Experiment 2 & 3 Nonexemplar Results

Experiment 2 Results

Reaction Time Analysis. *Prime x sPIP Analysis.* The additive model was favoured over the fully interactive model, *BF* = 68.53, θ = 2.16. None of the main effects or interactions were significant in this analysis, all *t*s < 1.28, all *p*s > .20.

Prime x iPIP Analysis. Once again, the additive model was favoured over the full model, *BF* = 55.24, θ = 1.74. The only effect that was significant in this analysis was the effect of iPIP, *β* = -9.14, *SE* = 4.53, *t*(7416) = -2.02, *p* = .044.

Prime x PIP Analysis. As with the sPIP and iPIP analyses, the additive model was favoured over the full model, *BF* = 36.05, θ = 1.14. None of the main effects, or the interaction were significant in any of the analyses, all *t*s < 1.55, all *p*s > .10.

Error Analysis. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 50.37, θ = 1830824462, but none of the effects in this analysis were significant, all *t*s < 1.55, all *p*s > .12. Overall, targets that were preceded by translation primes (*M* = 3.27%) produced similar error rates as targets preceded by control primes (*M* = 3.51%). Further, subjects in Tertile 1 (*M* = 3.57%), Tertile 2 (*M* = 3.19%), and Tertile 3 (*M* = 3.42%) produced the same error rates. Finally, there was no difference in the effect of primes on error rates in any of the tertiles.

Prime x iPIP Analysis. The additive model was favoured over the interactive model, *BF* = 63.88, θ = 2384754990, but none of the effects were significant, *t*s < 1. Targets in Tertile 1 (*M* = 3.17%), Tertile 2 (*M* = 3.73%), and Tertile 3 (*M* = 3.27%) produced similar error rates, and there was no difference in the effect of primes on error rates in any of the tertiles.

Prime x PIP Analysis. The additive model was favoured over the interactive model again, *BF* = 64.95, θ = 2360406057, but there were once again no significant effects in any of the analyses, *t*s < 1. Error rates in Tertile 1 (*M* = 3.69%), Tertile 2 (*M* = 2.60%), and Tertile 3 (*M* = 3.49%) were not significantly different, and there was no difference in the effect of the prime on error rates in any of the tertiles.

Experiment 3 Results

Reaction Time Analysis Using Experiment 2 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the fully interactive model, *BF* = 46.87, θ = 2.61, but none of the analyses involved any significant effects, *t*s < 1.40, *p*s > .16.

Prime x iPIP Analysis. The additive model was again favoured over the interactive model, *BF* = 23.14, θ = 1.29, but none of the analyses involved any significant effects, *t*s < 1.45, *p*s > .14.

Prime x PIP Analysis. The additive model was favoured over the interactive model in the Bayes Factor analysis, *BF* = 1.96, but not in the relative likelihood analysis, θ = 0.11. None of the main effects or interactions were significant in any of the models, *t*s < 1.40, *p*s > .16.

Error Analysis Using Experiment 2 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 49.06, θ = 2.63. None of the effects were significant in any of the analyses, *z*s < 1.

Prime x iPIP Analysis. The additive model was favoured over the interactive model, *BF* = 50.47, θ = 2.70. Again, none of the effects were significant in any analysis, *z*s < 1.

Prime x PIP Analysis. The additive model was once again favoured over the interactive model, *BF* = 49.87, θ = 2.67. None of the effects were significant in any analysis, *z*s < 1.

Reaction Time Analysis Using Experiment 3 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 48.78, θ = 2.72. None of the effects were significant in any analysis, *t*s < 1.45, *p*s > .14.

Prime x iPIP Analysis. The additive model was favoured over the interactive model, *BF* = 48.22, θ = 2.69. None of the effects were significant in any of the analyses, *t*s < 1.38, *p*s > .15.

Prime x PIP Analysis. The additive model was favoured over the interactive model, *BF* = 48.35, θ = 2.70. None of the effects were significant in any of the analyses, *t*s < 1.35, *p*s > .17.

Error Analysis Using Experiment 3 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 40.72, θ = 2.18. None of the effects were significant in any analysis, *z*s <1.

Prime x iPIP Analysis. The additive model was favoured over the interactive model, *BF* = 47.86, θ = 2.56. None of the effects were significant in any analysis, *z*s < 1.

Prime x PIP Analysis. Once again, the additive model was favoured over the interactive model, *BF* = 39.94, θ = 2.14, and none of the effects were significant in any of the analyses, *z*s < 1.

Combined Results of Experiments 2 and 3

Reaction Time Analysis Using Experiment 2 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 77.53, θ = 2.13. None of the effects were significant, *t*s < 1.

Prime x iPIP Analysis. The additive model was favoured over the interactive model, *BF* = 99.47, θ = 2.73. None of the effects were again significant, *t*s < 1.

Prime x PIP Analysis. The additive model was once again favoured over the interactive model, *BF* = 76.80, θ = 2.11. The only effect that was significant in this analysis was the effect of PIP, β = -10.71, *SE* = 5.42, $t(9794) = -1.98$, $p = .048$. Overall, larger PIP scores were associated with faster RTs than lower PIP scores.

Error Analysis Using Experiment 2 Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 47.34, $\theta = 1.26$, but none of the effects were significant, *z*s < 1.27, *p*s > .20.

Prime x iPIP Analysis. The additive model was again favoured over the interactive model, *BF* = 101.00, $\theta = 2.69$, and none of the effects were again significant, *z*s < 1.

Prime x PIP Analysis. The additive model was again favoured over the interactive model *BF* = 69.17, $\theta = 1.84$, and none of the effects were significant in any analysis, *z*s < 1.

Reaction Time Analysis Using Combined Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 96.98, $\theta = 2.66$, but none of the effects were significant in any analysis, *t*s < 1.

Prime x iPIP Analysis. The additive model was again favoured over the interactive model, *BF* = 93.58, $\theta = 2.57$, but again, none of the effects were significant in any analysis, *t*s < 1, *t*s < 1.22, *p*s > .21.

Prime x PIP Analysis. The additive model was again favoured over the interactive model, *BF* = 98.75, $\theta = 2.71$. Again, none of the effects were significant in any analysis, *t*s < 1.42, *p*s > .15.

Error Analysis Using Combined Coefficients. *Prime x sPIP Analysis.* The additive model was favoured over the interactive model, *BF* = 50.01, $\theta = 1.33$, but none of the effects were significant in any analysis, *z*s < 1.35, *p*s > .17.

Prime x iPIP Analysis. The additive model was again favoured over the interactive model, *BF* = 100.76, $\theta = 2.69$, but none of the effects were again significant, *z*s < 1.49, *p*s > .13.

Prime x PIP Analysis. The additive model was again favoured over the interactive model, *BF* = 77.32, θ = 2.06, but again, none of the effects were significant, zs < 1.

Appendix G

Ethics Applications for Data Collection

Western Research

Research Ethics

Western University Non-Medical Research Ethics Board
NMREB Delegated Initial Approval Notice

Principal Investigator: Prof. Stephen Lupker
Department & Institution: Social Science\Psychology,Western University

NMREB File Number: 108835
Study Title: An examination of lexical processing in Chinese-English bilinguals

NMREB Initial Approval Date: February 08, 2017
NMREB Expiry Date: February 08, 2018

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Other	Language Questionnaire	2017/02/06
Recruitment Items	SONA Description	2017/02/06
Letter of Information & Consent		2017/02/06
Western University Protocol	Received January 23, 2017	

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the above named study, as of the NMREB Initial Approval Date noted above.

NMREB approval for this study remains valid until the NMREB Expiry Date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario.

Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The NMREB is registered with the U.S. Department of Health & Human Services under the IRB [redacted]

Western Research

Date: 23 January 2018

To: Prof. Stephen Lupker

Project ID: 108835

Study Title: An examination of lexical processing in Chinese-English bilinguals

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Date Approval Issued: 23/Jan/2018

REB Approval Expiry Date: 08/Feb/2019

Dear Prof. Stephen Lupker,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

[redacted]

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Curriculum Vitae

Mark McPhedran

EDUCATION

Ph.D. Psychology, Cognition & Perception
University of Western Ontario (UWO)
September 2014 – December 2018

MSc. Psychology, Cognition & Perception
University of Western Ontario (UWO)
September 2012 – August 2014
Masters Thesis: *The Effects of Semantic Neighborhood Density on the Processing of Ambiguous Words*
Supervisor: Dr. Stephen Lupker

B.A. Psychology, Honours with Thesis
University of Windsor
September 2007 – June 2011
Honours Thesis: *Effects of Semantic Neighborhood Density on the Processing of Ambiguous and Unambiguous Words*
Supervisor: Dr. Lori Buchanan

EMPLOYMENT HISTORY

Lab Instructor and Teaching Assistantship: Psychology 2800E, Introduction to Research Methods in Psychology, University of Western Ontario, Sep 2017-Apr 2018.
Teaching Assistantship: Psychology 3139B, Cognitive Science, University of Western Ontario, Jan 2017-Apr 2017.
Teaching Assistantship: Psychology 1000, Introduction to Psychology, University of Western Ontario, Sep 2015-Dec 2016.
Teaching Assistantship: Psychology 2040A, Child Development, University of Western Ontario, Sep 2014-Dec 2014.
Teaching Assistantship: Psychology 3441F, Frontal Cortex and the Development of Cognitive Control, University of Western Ontario, Sep 2014-Dec 2014.
Teaching Assistantship: Psychology 3184G, Research Methods in Psycholinguistics, University of Western Ontario, Jan 2014-Apr 2014.
Teaching Assistantship: Psychology 1000, Introduction to Psychology, University of Western Ontario, May 2013-Jul 2013.
Teaching Assistantship: Psychology 2135B, Cognitive Psychology, University of Western Ontario, Jan 2013-Apr 2013.
Teaching Assistantship: Psychology 2135A, Cognitive Psychology, University of Western Ontario, Sep 2012-Dec 2012.

VOLUNTEER EXPERIENCE

Graduate Student Issues Committee (GSIC)
Committee Member
November 2015 – July 2018
The Graduate Student Issues Committee is involved with general advocacy initiatives in addressing issues that graduate students face on campus, and actively campaigning for graduate student interests.

The Grad Club Committee
Committee Member
November 2015 – July 2018
I was on the governing board for the Grad Club at Western University. This volunteering involves attending monthly meetings and reviewing issues that the Grad Club is facing, such as budgeting, menu selection, and any other administrative work that is required for the upkeep of the Grad Club.

CONFERENCE PRESENTATIONS

McPhedran, M.J., & Lupker, S.J. (June 2017). On the joint effects of stimulus quality and masked repetition priming. Presented at the Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Sciences (CSBBCS), University of Regina, Regina, SK.

McPhedran, M., Taikh, A., Spinelli, G., & Lupker, S.J. (November 2016). The impact of visible intervenors on form and identity priming. Presented at the Annual Meeting of the Psychonomic Society, Boston, MA.

McPhedran, M.J., (March 2016). Effects of stimulus quality on masked repetition priming in lexical decision: The pitfalls of contrast reduction in masked priming. Presented at the Annual Western Research Forum, London, ON.

CERTIFICATIONS

Western Certificate in University Teaching and Learning, October 2018.
Teaching Assistant Training Program Certificate, September 2012.

RESEARCH INTERESTS

- Semantic memory and lexical processing.
- Visual word recognition, reading, and general psycholinguistic processes.
- Bilingualism and cognitive development.
- Applications of big data, machine learning, computational modeling, and natural language processing to the behavioural sciences.
- Cognitive control and executive processes.