

**BIDIRECTIONAL TRANSFER: CONSEQUENCES OF TRANSLATION AMBIGUITY  
FOR BILINGUAL WORD MEANING**

by

**Tamar Degani**

B.A., University of Haifa, Israel, 2003

M.S., University of Pittsburgh, 2007

Submitted to the Graduate Faculty of  
Arts and Sciences in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Tamar Degani

It was defended on

June 14<sup>th</sup>, 2011

and approved by

Charles A. Perfetti, Professor, Department of Psychology

David C. Plaut, Professor, Department of Psychology, Carnegie Mellon University

Tessa Warren, Associate Professor, Department of Psychology

Dissertation Advisor: Natasha Tokowicz, Assistant Professor, Department of Psychology

Copyright © by Tamar Degani

2011

# **BIDIRECTIONAL TRANSFER: CONSEQUENCES OF TRANSLATION**

## **AMBIGUITY FOR BILINGUAL WORD MEANING**

Tamar Degani, PhD

University of Pittsburgh, 2011

Could a second language (L2) influence how bilinguals process their native language (L1)? The work described in this dissertation examined this issue focusing on the way bilinguals interpret the meanings of words. Capitalizing on the prevalence of words that can be translated in more than one way across languages (i.e., *translation ambiguity*, e.g., ‘watch’ and ‘clock’ are both translated into Spanish as ‘reloj’), the current work examined if and how bilinguals are affected by the indirect mapping between translations. Performance of two groups of bilinguals who differed in the order in which they learned English and Spanish (English-Spanish and Spanish-English bilinguals) was compared to that of monolingual English controls. In Experiment 1 participants’ eye movements were recorded as they read English sentences, in which target words were replaced with English words that either share a translation with the target in Spanish (e.g., ‘clock’ to replace ‘watch’, both corresponding to ‘reloj’ in Spanish) or a control. Participants’ sensitivity to the degree of anomaly created by these replacements was compared. The results indicate bidirectional patterns of transfer (L1 influence on L2 and the reverse), in that both bilingual groups processed shared-translation replacements differentially than controls. Experiment 2 further asked if translation ambiguity impacts intra-word senses. Participants were presented with pairs of phrases instantiating different senses of ambiguous English words (e.g., dinner date – expiration date) and were asked to decide if the two senses were related in meaning. Critically, for some pairs of phrases a single Spanish word encompassed both

meanings of the ambiguous word ('joint-translation') and for others each sense corresponded to a different Spanish translation ('split-translation'). The proportion of 'yes' responses and latency data again suggest differences between bilinguals and monolinguals as a function of translation status in Spanish. These results demonstrate that language experience continues to shape semantic representations, and highlight the dynamic and interconnected nature of the bilingual lexicon. L2 learning can thus impact the meaning interpretation of words, and may lead to subtle differences in semantic processing between monolingual and bilingual speakers.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>XIII</b>
<b>1.0 GENERAL INTRODUCTION.....</b>	<b>1</b>
<b>1.1 BILINGUALISM AND BIDIRECTIONAL TRANSFER.....</b>	<b>3</b>
<b>1.2 TRANSLATION AMBIGUITY.....</b>	<b>6</b>
<b>1.3 OVERVIEW OF THE CURRENT RESEARCH.....</b>	<b>13</b>
<b>2.0 EXPERIMENT 1: SHARED TRANSLATION EFFECT IN SENTENCES USING EYE TRACKING.....</b>	<b>18</b>
<b>2.1 INTRODUCTION .....</b>	<b>18</b>
<b>2.1.1 The shared-translation effect .....</b>	<b>20</b>
<b>2.1.2 Semantic anomaly in eye tracking studies .....</b>	<b>29</b>
<b>2.1.3 The present experiment.....</b>	<b>32</b>
<b>2.2 METHOD .....</b>	<b>35</b>
<b>2.2.1 Participants.....</b>	<b>35</b>
<b>2.2.2 Materials .....</b>	<b>38</b>
<b>2.2.3 Procedure.....</b>	<b>40</b>
<b>2.2.3.1 General procedure .....</b>	<b>40</b>
<b>2.2.3.2 Sentence reading. ....</b>	<b>41</b>
<b>2.2.3.3 Sentence naturalness rating.....</b>	<b>42</b>

<b>2.3</b>	<b>RESULTS .....</b>	<b>43</b>
2.3.1	Sentence reading task .....	43
2.3.1.1	Data pre-processing and trimming .....	43
2.3.1.2	Dependent measures .....	44
2.3.1.3	Analysis approach.....	45
2.3.1.4	Target region.....	47
2.3.1.5	Post-target region.....	53
2.3.1.6	Summary of main results across regions .....	60
2.3.1.7	Summary of covariates' effects across regions .....	65
2.3.1.8	Group differences in reading natural (IST) sentences .....	66
2.3.2	Sentence naturalness ratings.....	68
2.3.2.1	Group differences in ratings of natural (IST) sentences .....	71
<b>2.4</b>	<b>DISCUSSION.....</b>	<b>72</b>
<b>3.0</b>	<b>EXPERIMENT 2: THE SPLIT-TRANSLATION EFFECT IN JUDGMENTS OF SEMANTIC RELATEDNESS.....</b>	<b>83</b>
<b>3.1</b>	<b>INTRODUCTION .....</b>	<b>83</b>
3.1.1	Semantic relatedness effects in ambiguity processing .....	86
3.1.2	Inhibition and facilitation among intra-word senses.....	91
3.1.3	The current experiment.....	95
<b>3.2</b>	<b>METHOD .....</b>	<b>99</b>
3.2.1	Participants.....	99
3.2.2	Materials .....	100
3.2.3	Procedure.....	101

<b>3.3</b>	<b>RESULTS .....</b>	<b>102</b>
3.3.1	Data analyses .....	102
3.3.2	Response probability .....	106
3.3.3	Reaction times .....	108
<b>3.4</b>	<b>DISCUSSION.....</b>	<b>110</b>
3.4.1	Language proficiency.....	113
3.4.2	Location of the modifiers.....	116
3.4.3	Baseline semantic relatedness .....	117
<b>4.0</b>	<b>GENERAL DISCUSSION .....</b>	<b>120</b>
	<b>APPENDIX A - STIMULI CREATION AND NORMING FOR EXPERIMENT 1 .....</b>	<b>124</b>
	<b>APPENDIX B - CRITICAL WORD TRIPLETS FOR EXPERIMENT 1.....</b>	<b>138</b>
	<b>APPENDIX C - CRITICAL EXPERIMENTAL SENTENCES EXPERIMENT 1 .....</b>	<b>142</b>
	<b>APPENDIX D - METHODS AND RESULTS FOR PROFICIENCY AND INDIVIDUAL DIFFERENCES TASKS .....</b>	<b>151</b>
	<b>APPENDIX E - ANALYSES OF THE PRE-TARGET REGION EXPERIMENT 1 .....</b>	<b>164</b>
	<b>APPENDIX F - RELATEDNESS RATINGS OF EXPRESSION PAIRS FOR EXPERIMENT 2 .....</b>	<b>170</b>
	<b>APPENDIX G - CRITICAL EXPRESSION PAIRS EXPERIMENT 2.....</b>	<b>173</b>
	<b>APPENDIX H - TASK INSTRUCTIONS .....</b>	<b>181</b>
	<b>APPENDIX I - LANGUAGE HISTORY QUESTIONNAIRE .....</b>	<b>187</b>
	<b>APPENDIX J - VOCABULARY POST-TEST .....</b>	<b>191</b>
	<b>BIBLIOGRAPHY .....</b>	<b>195</b>

## LIST OF TABLES

Table 1. Background characteristics for the final set of participants by group .....	37
Table 2. Experiment 1 example sentences.....	38
Table 3. Experiment 1 stimuli characteristics.....	39
Table 4. Coefficient estimates for the target region.....	47
Table 5. Model estimates for FFD, GD, and GPT in the target region as a function of group and semantic relatedness.....	50
Table 6. Model estimates for TT in the target region as a function of condition and semantic relatedness.....	50
Table 7. Coefficient estimates for the post-target region.....	53
Table 8. Mean (and SD) performance on IST sentences in each sentence region and measure as a function of group.....	67
Table 9. Coefficient estimates for the sentence naturalness rating task. ....	69
Table 10. Experiment 2 example stimuli and characteristics by condition. ....	101
Table 11. Coefficient estimates for semantic relatedness judgments. ....	104
Table 12. Coefficient estimates for semantic relatedness judgments including relatedness interactions.....	105
Table 13. Percentage of ‘yes’ responses as a function of group and relatedness. ....	107

Table 14. Background characteristics for the participants in norming experiments.....	126
Table 15. Critical word triplets for Experiment 1 .....	138
Table 16. Critical experimental sentences Experiment 1.....	142
Table 17. Rotated component matrix.....	162
Table 18. Coefficient estimates for the pre-target region. ....	164
Table 19. Model estimates for mean performance in the pre-target region as a function of group, condition, and semantic relatedness.....	167
Table 20. Coefficient estimates for the pre-target region excluding trials with target skips .....	168
Table 21. Critical expression pairs Experiment 2.....	173

## LIST OF FIGURES

Figure 1. Schematic representation of the shared-translation and split-translation effects in the mapping of .....	15
Figure 2. Skipping probability in the target region as a function of group, condition, and semantic relatedness.....	48
Figure 3. Differences in skipping probability in the target region between DTR and STR sentences as a function of group and semantic relatedness. ....	49
Figure 4. Regression-out probability from the target region as a function of group and condition .....	51
Figure 5. Skipping probability in the post-target region as a function of group, condition, and semantic relatedness.....	54
Figure 6. Differences in skipping probability in the post-target region between DTR and STR sentences as a function of group and semantic relatedness. ....	55
Figure 7. Go-past time in the post-target region as a function of group, condition, and semantic relatedness.....	56
Figure 8. Differences in go-past time in the post-target region between DTR and STR sentences as a function of group and semantic relatedness.....	56

Figure 9. Total viewing time in the post-target region as a function of group, condition, and semantic relatedness.....	57
Figure 10. Differences in total viewing time in the post-target region between DTR and STR sentences as a function of group and semantic relatedness. ....	58
Figure 11. Regression-out probability from the post-target region as a function of group, condition, and semantic relatedness.....	59
Figure 12. Differences in regression-out probability from the post-target region between DTR and STR sentences as a function of group and semantic relatedness. ....	59
Figure 13. Mean sentence naturalness ratings as a function of group, condition, and semantic relatedness.....	70
Figure 14. Mean difference in sentence naturalness ratings between DTR and STR sentences as a function of group and semantic relatedness.....	71
Figure 15. Proportion of ‘yes’ responses as a function of condition and group. ....	107
Figure 16. Response times for ‘no’ responses as a function of condition and group. ....	109
Figure 17. Proportion of ‘yes’ responses of SE bilinguals as a function of condition and English proficiency. ....	115
Figure 18. Response times for ‘no’ responses of SE bilinguals as a function of condition and English proficiency. ....	115
Figure 19. Skipping probability of the pre-target region as a function of group, relatedness, and condition. ....	166
Figure 20. Skipping probability of the pre-target region as a function of group.....	166

## ACKNOWLEDGEMENTS

Amazing people have helped me complete this dissertation project. I am thankful to my advisor, Natasha Tokowicz, for giving me the tools, guidance and confidence to complete my work, and for truly accepting my choices throughout the way. For allowing me to develop as a researcher and a mother at the same time, and for encouraging diversity in thought, style, and language. I am thankful to my committee members, Tessa Warren, Chuck Perfetti, and David Plaut, for their insights and suggestions, for expressing interest and sparking ideas on various aspects related to this work. And thank you Tessa, for your support and encouragement throughout my graduate career. To Elizabeth Votruba-Drzal, for the solid statistical grounds you helped me develop, for your guidance through the complexities of the analyses described in this thesis, and mostly for your friendship and support throughout my graduate studies. To my friend Jessica Nelson, who made it possible for me to use the LMEs and for making ‘R dates’ a fun thing. And to Nikole Patson, for introducing me to the intricacies of eye movement research, but mostly for being my friend and making office visits such a pleasure. To the recent graduate and dear friend Leida Tolentino, for your friendship, support, and insights over the years, and for becoming my kids’ official ‘tia’. For knowing what priorities are, and for making our journey together such a pleasure. To my fellow Plum lab graduate students, Chelsea Eddington, Alba Tuninetti, and Alison Phillips, for all your help in creating and conducting this research, and for your excitement

and interest in the wonders of the bilingual mind. To my research assistants, especially Jessica Steiner and Kristin Schulmeister, who made it possible to collect eye movements from so many participants, and to all 400 participants who actually took part in the experiments described here. Couldn't have done it without you.

To Nuria Jane, for making Spanish accessible in countless dates of stimuli development, for your attention, your kindness, your suggestions and intuitions, and mostly for being my honest friend over the years. To my dear friend Laura Halderman, for being there to listen and think through whatever was going on, for lasagna in times of need. To Anat Prior, my eternal friend and conference roommate, who got excited with me on these questions from the beginning. For knowing to say the right words in the right time, for encouraging, supporting, and believing.

To my parents, my brother and my sister, for endless love and support, for going through everything with me, even from a distance, and for being patient enough to know it will have to be over at some point. And mostly, to my true partner in this journey, my husband Amir, for always listening, understanding, supporting, and encouraging, making it all possible. And lastly, to my amazing kids, Yuval and Eyal, who really make it all worthwhile.

## 1.0 GENERAL INTRODUCTION

Words gain their significance from the mapping they offer between lexical form and meaning. Critically, languages may differ not only in the lexical forms they use, but also in the mappings between forms and meanings. As will be described below, these differences lead to cases in which a given word form may correspond to more than one translation in another language (i.e., *translation ambiguity*, see below). Moreover, each language is filled with ambiguities in the mapping of words to meanings, such that a given meaning can sometimes be expressed by more than one lexical form creating synonyms, and one lexical form may encompass more than one meaning or sense creating semantic ambiguity. These indirect mappings of meanings and forms within and across languages may bear consequences for how speakers of different languages, and importantly multilingual speakers, process the meaning of words.

The goal of the work described in this thesis is therefore to explore the consequences of ambiguity within and across languages for bilinguals' meaning representations. We specifically examine if the mapping of words to meanings in the first language (L1) of bilingual speakers can influence processing of word meaning in the L2. Critically, we also examine if a later-learned L2 can similarly influence processing of L1 word meaning. As will be described below, such bidirectional cross-language influences would demonstrate the interconnectivity between the languages of bilingual speakers and the dynamic nature of the mental lexicon. Moreover, if speakers process words partially in accordance with the mapping of words to meanings in their

other languages, important differences are likely to emerge between bilinguals and monolinguals.

Before comparing monolingual and bilingual speakers, it is important to unpack this classification. In particular, the term bilingualism is not restricted to people who learned both languages simultaneously or that achieved native-like proficiency in both languages. Rather, bilinguals are a heterogeneous group of individuals who use more than one language on a daily basis. This heterogeneity could be characterized along central dimensions which include age of acquisition, learning situation (immersion/classroom), proficiency and use (in writing/reading as well as talking/listening). In the current study we focus on bilingual individuals who learned one language after the other (sequential bilinguals), and who achieved at least moderate proficiency in their L2. One group of bilinguals is immersed in their L1 at the time of testing, and the other is immersed in their L2. Moreover, the monolingual speakers who serve as control have likely been exposed to languages other than English, because the requirement posed by many schools to learn an L2 makes it difficult to recruit monolingual English speakers without any exposure to other languages (for discussion of the difficulty in identifying monolinguals for a comparison group, see Cook, 2003). Nonetheless, monolinguals and bilinguals differ dramatically in their life-long language proficiency and use.

Before we describe the experiments that constitute this thesis, we first describe the relevant literatures on cross-language influences in the bilingual lexicon and the translation ambiguity that exists across languages.

## 1.1 BILINGUALISM AND BIDIRECTIONAL TRANSFER

Understanding the interplay between languages in the minds of multilingual speakers is relevant on both practical and theoretical grounds. First, the majority of the world's population uses more than one language in daily life (e.g., Edwards, 2004; for discussion, see also Cook, 2003) and recent initiatives have been established to promote L2 learning (e.g., the National Security Language Initiative in the United States), thus leading even more people to use multiple languages. Moreover, the study of bilingualism allows one to test the ways in which experience shapes the cognitive system. On the one hand, some theories hold that early experience alone can guide cognitive organization, such that experiences that are received after a certain critical period carry little, if any, consequences for the organization and performance of the cognitive system. In the case of language, such critical period theories (Lenneberg, 1967; see also DeKeyser & Larson-Hall, 2005) hold that the L1 is privileged, not only in being easier to learn, but also in its influence on representation and processing. Any languages learned later in life are bound to rely on the L1. Data suggesting that bilinguals still show traces of their native language when they process their L2 has been taken to suggest the irrevocable influence of the L1 (e.g., Jiang, 2002).

In contrast, other theories posit that the cognitive system continues to change as new experiences are encountered, and that although early experiences may carry more weight, new experiences can nonetheless exert an influence on cognitive representation and organization (e.g., MacWhinney, 2005). In the case of language this will be reflected in influences of a later-learned L2 on L1 representation and processing. In recent years, this notion of bidirectional influences, according to which the L2 may also influence the native language, has begun to gain support (e.g., Ameel, Storms, Malt, & Sloman, 2005; Ameel, Malt, Storms, & van Assche, 2009;

Brown, & Gullberg, 2008; Cook, 2003; Dong, Gui, & MacWhinney, 2005; Hohenstein, Eisenberg, & Naigles, 2006; Pavlenko & Jarvis, 2002; Wolff & Ventura, 2009).

For example, in an oral-narrative analysis of productions from Russian-English bilinguals, Pavlenko and Jarvis (2002) observed instances of bidirectional transfer, from L1 to L2 and importantly from L2 to L1, in both semantic and syntactic aspects. For instance, word choice was influenced by both languages, such that bilinguals either borrowed a lexical term from their other language, or extended the meaning of a word based on their other language (as in using the word 'neighbor' to refer to roommate, because the Russian translation 'sosed' encompasses both). Such semantic extensions are of particular relevance here, because they demonstrate a potential bidirectional influence of translation ambiguity. Interestingly, Pavlenko and Jarvis examined the influence of factors such as participants' length of immersion and age of arrival, as well as the context and mode within which the productions were elicited, but observed no significant influence of these external factors on the amount and direction of the transfer. These results suggest that even late learners of L2 can, and do, exhibit influences of the L2 on L1. They further highlight the need to examine these bidirectional transfer patterns, rather than assume that only the L1 can influence processing in L2.

Similarly, in narrative analyses of autobiographical memories, Marian and Kaushanskaya (2007) observed instances in which bilinguals mistakenly used a word that shares a translation with the intended word in their other language (as in 'at the start' instead of 'at the beginning' because a single Russian word 'nachalo' refers to both). Critically, here too, the authors identified instances of transfer from both L1 to L2 and the reverse.

A continued influence of L1 on L2 results in non-native performance of L2 learners in their L2. However, bidirectional cross-language influences, in which the L2 also influences L1

processing, results in differences between bilinguals and monolinguals in *both* of their languages. Indeed, Ameel et al. (2005; Ameel et al., 2009) showed that when asked to name simple household objects, bilinguals exhibited naming patterns in both Dutch and French that differ from those of monolingual speakers of those languages. Bilinguals' naming patterns yielded evidence for convergence, such that naming in L1 and L2 was more similar to each other in bilinguals than were the two naming patterns of monolinguals in each language. This suggests an influence in both languages from the word-to-referent mappings in the other language. Wolff and Ventura (2009) similarly showed that Russian-English bilinguals' expressions of causation exhibit cross-language influences that make their expressions different from those of both monolingual Russian and monolingual English speakers.

These findings suggest that cross-language influences are possible in both directions. Although not explicitly captured by models of the bilingual lexicon, these influences are expected based on models that assume interconnectivity and shared semantic representations. These models (e.g., *Revised Hierarchical Model*, Kroll & Stewart, 1994; *Distributed Feature Model*, de Groot, 1992; *Distributed Conceptual Feature Model*, van Hell & de Groot, 1998) provide the foundation for bidirectional influences to emerge. Because words in the two languages are interconnected and both access (albeit with different ease) a shared semantic level, L2 processing is expected to be influenced by the mapping of words to meanings in L1, as well as the reverse. Interconnectivity between words in the two languages of bilinguals is also a key assumption in bilingual models of lexical access such as the Bilingual Interactive Activation models (BIA and BIA+, Dijkstra & Van Heuven, 1998; 2002). These models maintain that inter-language and intra-language connections are both present within an integrated lexicon, which includes words from both L1 and L2. Activation flow could therefore lead to associations

between words in L1 and in L2. A bidirectional manifestation of the cross-language influences of translation ambiguity of the type examined in the current study would support notions of the bilingual lexicon as dynamic and interconnected in nature.

A central goal of the current work is therefore to examine cross-language influences in both directions. We focus on bilinguals for which L1 was clearly learned prior to the L2, to allow examination of the direction of the effects. We test the influence of Spanish on English in two groups of bilinguals who differ in the order in which they learned their languages. Spanish-English bilinguals will potentially exhibit influences of L1 (Spanish) on L2 (English), and English-Spanish bilinguals will potentially exhibit influences of L2 (Spanish) on L1 (English). We make use of translation ambiguity to examine these bidirectional influences.

## 1.2 TRANSLATION AMBIGUITY

Words may have more than one translation across languages. This phenomenon, which is termed *translation ambiguity* (e.g., Prior, MacWhinney, & Kroll, 2007; Tokowicz & Kroll, 2007; Tokowicz, Kroll, de Groot, & van Hell, 2002), creates numerous cases in which two or more words correspond to a shared translation in the other language. For instance, the English words 'fault' and 'guilt' correspond to the same Spanish word 'culpa'. Importantly, it is not the case that only a few exception words correspond to more than one translation in another language. Rather, as has been demonstrated empirically in recent years (e.g., Eddington, Degani, & Tokowicz, 2011; Prior et al., 2007; Tokowicz et al., 2002), translation ambiguity is a common phenomenon in the mapping between languages. Tokowicz et al. (2002) demonstrated this when they showed that approximately 25% of English/Dutch translations that were assumed to have only one

translation by previous research, in fact elicited more than one translation when bilinguals were asked to translate them across languages.

Using the same sample of English words, Eddington et al. (2011) showed that over 40% of the words elicited multiple translations in German. Furthermore, words that tended to elicit multiple translations from English to Dutch, were also more likely to elicit multiple translations from English to German. This suggests that ambiguity in the source language (English in this case) is at least partially responsible for the existence of multiple translations, and that synonymy in the target language (German or Dutch in this case) could not solely account for translation ambiguity.

In fact, there are multiple sources that could contribute to the presence of translation ambiguity (see e.g., Prior, Wintner, MacWhinney, & Lavie, 2011, for review). As mentioned above, near-synonymy in the target language can lead to the existence of more than one lexical choice when translating a given word (e.g., couch and sofa are both correct translations of the Spanish word ‘sofá’). Semantic ambiguity, in the form of homonymy (e.g., bark) and polysemy (e.g., line), further leads to multiple translations, because each meaning/sense of the ambiguous word is likely to receive a separate translation in another language (for extended discussion, see Chapter 3). Additionally, part-of-speech ambiguity and morphological ambiguity within a language are also likely to result in multiple translations in another language, such that separate translations denote different word classes or morphological derivations of the source word (e.g., ‘cook’ is translated as ‘cocinero’ in Spanish to denote the noun meaning, and ‘cocinar’ to denote the verb meaning of the word; see Prior et al., 2007).

Of particular relevance to the current study, translation ambiguity is especially prevalent between English and Spanish. This is partly because both languages are widely spoken, resulting

in potential local variations in word use (Prior et al., 2011). Prior et al. (2007) found that about 60% of Spanish/English translations elicited more than one translation from bilingual speakers. In that sample, special attention was given to part-of-speech ambiguity, which is prevalent in English. Word class ambiguous items (e.g., cook) elicited more translations into Spanish compared to verbs, which in turn resulted in more translations than nouns. Further, the probability of each translation was computed across participants, and was related to word characteristics and to the form similarity of the translations. Specifically, translations that were higher in frequency and imageability, and that were more similar in form to the cue word, were more likely to be given as translations. Degani, Eddington, Tokowicz, and Prior (2009) further showed that in addition to lexical characteristics and form similarity, meaning probability within a language was a significant predictor of translation probability, suggesting that ambiguity in the source language is relevant to translation ambiguity (see also Eddington et al., 2011).

Interestingly, the controlled laboratory findings of translation probability in isolation received further support as reflecting bilinguals' life-long experience with the distributions of the alternative translations (Prior et al., 2011). In particular, Prior et al. (2011) found moderate correlations between translation probabilities from the lab (Prior et al., 2007) and those obtained from large parallel language corpora, in which professional translators translated the words in a meaningful context. Furthermore, lexical characteristics (frequency and imageability) seem to be more influential in determining translation choice in context, but form similarity of the translation to the cue word is more influential in the decontextualized lab task.

Together, these studies demonstrate that translation ambiguity is widespread both in isolation and in context, and that it is linked to ambiguity within a language. Translation ambiguity has further been shown to influence performance of bilinguals and L2 learners (see

Tokowicz & Degani, 2010, for a review). Degani and Tokowicz (2010a) demonstrated that translation ambiguous words were more difficult to learn than translation unambiguous words. In a multiple-session training study, native English speakers learned Dutch vocabulary that could either map in a one-to-one fashion to English translations, or could be translation-ambiguous such that two Dutch words corresponded to a single English word. This ambiguity was either due to synonymy in Dutch (e.g., both 'lucht' and 'hemel' denote 'sky' in English), or to semantic ambiguity in English (e.g., 'change') such that each Dutch word encompassed a different meaning of the ambiguous English word (e.g., 'verandering' to denote alteration, and 'wisselgeld' to denote coins). In both translation production and translation recognition tests, Degani and Tokowicz observed better performance for translation-unambiguous words, and this advantage was especially strong in comparison to translation ambiguous words that mapped onto a single meaning in English (i.e., synonymous Dutch words). The authors suggested that the difficulty associated with learning translation-ambiguous words is due to the need to map multiple lexical forms to a single, undifferentiated meaning.

Translation ambiguity continues to exert its influence even for proficient bilinguals (Eddington & Tokowicz, 2011; Laxén & Lavaur, 2010; Michael, Tokowicz, Degani, & Smith, 2011; Tokowicz & Kroll, 2007). Tokowicz and Kroll (2007) showed that translation ambiguity impacted English-Spanish bilinguals' production performance. Translation production of words with multiple translations was less accurate than production of words with a single English-Spanish translation. Furthermore, the RT data revealed a disadvantage for words with multiple translations, which was driven by abstract words. Speed of production of concrete words, in contrast, was less affected by number of translations. The authors suggested that active

competition between multiple alternatives likely makes it more difficult to select one option for production.

However, subsequent research demonstrated that translation recognition performance, in which no single word is to be selected for response output, is still affected by the availability of multiple translations (Laxén & Lavuar, 2010; see also Boada, Sánchez-Casas, García-Albea, Gaviln, & Ferr, 2009). Laxén and Lavuar (2010) showed that French-English bilinguals were slower and less accurate to indicate that two words were translations of each other when another alternative translation existed. The disadvantage for translation-ambiguous words was stronger when the less probable (non-dominant) translation was presented, and was also stronger when the two alternative translations were unrelated in meaning. The translation ambiguity effect held when either the L1 or the L2 was presented first. In a primed translation recognition task, Eddington and Tokowicz (2011) similarly observed that English-German bilinguals recognized translation-ambiguous words less accurately and more slowly than translation-unambiguous words. Also, performance was faster and more accurate when the dominant translation was presented compared to the non-dominant translation.

The effect of translation ambiguity on bilingual performance seems to be modulated by individual differences in working memory and the ability to ignore task-irrelevant information (Michael et al., 2011). Specifically, intermediate learners of Spanish were less accurate in translating translation-ambiguous words compared to unambiguous words, but this difficulty was less pronounced for individuals with both higher working memory and better ability to ignore irrelevant information. Interestingly, the difficulty was *most* prominent for individuals with higher working memory and less ability to ignore task irrelevant information, presumably

because these individuals suffered greater interference from their knowledge of multiple alternative translations.

Importantly, because multiple sources give rise to translation ambiguity, the alternative translations can differ in their semantic relatedness. To illustrate, when an English homonym like the word 'bark' is translated into Spanish, its two translations ('ladrido' denoting the sound a dog makes and 'corteza' denoting the outer layer of a tree) are relatively unrelated in meaning. In contrast, translation ambiguity that results from synonymy (e.g., 'couch' and 'sofa' are both translations of the Spanish word 'sofá') creates a case in which the two alternative translations are, by definition, highly related in meaning. Critically, however, these are not categorically separated. Rather, in many cases the two alternative translations are moderately related, as when they correspond to a polysemous word (e.g., 'card' and 'letter' for 'carta' in Spanish). It is useful to capture this continuous nature of semantic relatedness. In a recent study we estimated the semantic relatedness of the different translations given to a source word as a continuous measure by computing what we call the Translation Semantic Variability (TSV) score. Specifically, based on translation norms, for each source (German) word, we averaged the semantic relatedness ratings of all possible pairings of the different English translations that were provided for that source word (Eddington et al., 2011; see also Eddington & Tokowicz, 2011). In the current study, we similarly sample translation ambiguity that falls along the continuum of semantic relatedness, such that the two translations of shared-translation words ranged in their baseline semantic relatedness.

In the current study, we examine the effect of such translation ambiguity on the way proficient bilinguals process the meaning of words. In particular, the 'shared-translation effect' refers to the possibility that two word forms that share a translation in another language may be

perceived differently by bilinguals than two word forms with different translations. As will be discussed in Chapter 2, the limited research available on this issue suggests a difference between shared-translation and different-translation pairs, but there are competing theories and contradictory evidence about the nature of this effect—i.e., whether shared-translation pairs are *more* or *less* similar in meaning than two word forms with different translations.

A second aspect of translation ambiguity, to be examined in Chapter 3, focuses on changes in the relatedness of two meanings/senses of a word due to translation ambiguity. In particular, we examine the possibility that when an ambiguous word (e.g., 'ring') has two translations in another language (e.g., 'anillo' and 'timbre' in Spanish, the first denoting jewelry and the second denoting the sound), bilinguals may perceive the word's senses as *less* similar, compared to an ambiguous word that has a single translation that captures all of its senses (e.g., the Spanish word 'cuerpo' captures both the biological and the administrative meanings of its English translation 'body'). The focus is on the meanings/senses of the ambiguous word itself, rather than on the translations.

The work described in this thesis seeks to identify possible influences of translation ambiguity on the relatedness of translation pairs and intra-word meanings/senses, and to test whether these translation-ambiguity effects are limited to the influence of L1 on L2, or whether bidirectional cross-language influences are observed. Whether the effects manifest themselves in a bidirectional fashion will provide important constraints for the theoretical explanations of the findings, and for the degree of interconnectivity postulated by models of the bilingual lexicon.

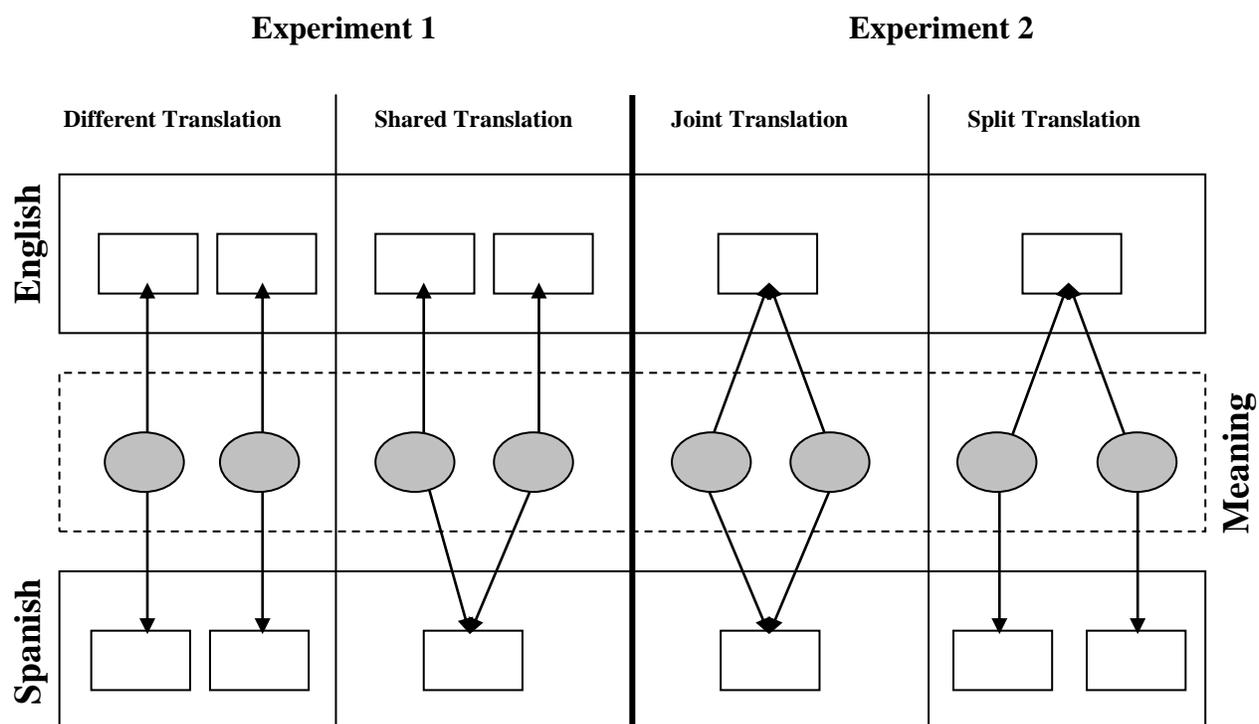
### 1.3 OVERVIEW OF THE CURRENT RESEARCH

The current study seeks to characterize the ways in which translation-ambiguity impacts bilingual word meaning, and further to test whether such influences are manifested in a bidirectional fashion between L1 and L2. In Experiment 1, we explore these bidirectional influences by focusing on cases in which a single word form in one language (Spanish) corresponds to two translations in another language (English). We test whether bilinguals are influenced by this one-to-many mapping. We ask if the English words 'candle' and 'sail', for instance, become more or less related for bilingual speakers who know Spanish, because the same Spanish word ('vela') is used to refer to both. We test both native Spanish speakers who learned English as an L2 (Spanish-English bilinguals) and native English speakers who learned Spanish as an L2 (English-Spanish bilinguals), such that across both groups we test the influence of L1 on L2 and vice versa. We further compare their performance to a group of monolingual English speakers. All participants are asked to read English sentences, in which on some proportion of trials target words are replaced with a word that shares a Spanish translation with the target (shared-translation replacement) or a control (different-translation replacement). For example, in the sentence 'To mask the smell of the burnt food, Alison lit a scented candle in the kitchen', the target word 'candle' is replaced with its shared-translation counterpart ('sail') or with a control word 'jail'. We track participants' eye movements to examine the degree to which such replacements cause disruptions in sentence comprehension (see Warren, 2011, for review).

In Experiment 2, we extend this paradigm and ask whether two meanings/senses of an ambiguous English word (e.g., body, as in administrative body vs. human body) become more or less related if a single lexical form corresponds to both meanings (e.g., 'cuerpo' in Spanish), compared to an ambiguous word for which separate Spanish translations denote each of its

senses/meanings (e.g., ‘anillo’ and ‘timbre’ denote different meanings of the ambiguous English word ‘ring’). The same three groups of participants (English-Spanish bilinguals, Spanish-English bilinguals, monolingual English speakers) are asked to make timed relatedness judgments to pairs of expressions instantiating different meaning/senses of ambiguous English words (e.g., diamond ring-loud ring, see Klepousniotou, Titone, & Romero, 2008 for similar stimuli but a different task). We examine if the proportion of ‘yes’ responses, as well as latencies to make yes and no responses, vary as a function of whether a single Spanish translation captures both senses of the ambiguous word (joint-translation condition) or two Spanish translations exist, each denoting a separate meaning of the ambiguous word (split-translation condition).

The logic guiding both experiments is similar, but as can be seen in Figure 1, each experiment taps slightly different components of the bilingual form-meaning system. In particular, in Experiment 1 we compare the left two columns of the figure, and in Experiment 2 we compare the right two columns of the figure. In Experiment 1 we estimate how interchangeable the two English words are when they map onto two Spanish translations (different translation) or a single Spanish translation (shared translation). Although measured as a distance between two English words, it presumably reflects the distance at the semantic level (denoted by circles in the figure). In Experiment 2 we attempt a more direct measure of the distance at the semantic level, by focusing on two meanings/senses of ambiguous English words. We measure the relatedness of two expressions denoting these two meanings. Thus in both cases we attempt to examine the change in relatedness (or distance) between the two meanings (denoted in the figure by circles), but we estimate those based on the English words (top squares in the figure).



**Figure 1.** Schematic representation of the shared-translation and split-translation effects in the mapping of words to meanings in the bilingual lexicon.

A few notes are important to make with respect to Figure 1. First, the schematic circles used to denote meaning are not to be taken as a statement about the nature of representation. These meanings could be composed of sets of semantic features, which may overlap to different degrees in different cases (e.g., van Hell & de Groot, 1998). Furthermore, the debate surrounding whether the different senses/meanings of ambiguous words are represented together or separately (e.g., Klein & Murphy, 2001; see Chapter 3) is outside the scope of the current work. Second, the figure does not capture the variability in semantic relatedness examined in the current study. In particular, within each column, words and meanings may be more or less related, irrespective of the mapping to the other language. Thus, ‘home’ and ‘house’ are more related than ‘candle’ and ‘sail’, although both map onto a single Spanish translation (‘casa’ and ‘vela’, respectively).

Similarly, the administrative meaning and biological meaning of the word ‘body’ are less related than the two senses of the word ‘cotton’ (cotton thread and cotton dress) although both are captured by a joint Spanish translation (joint translation condition). Third, because we test two groups of bilinguals that differ in the order in which they learned English and Spanish, lexical forms in English (top row of the figure) serve as L1 to English-Spanish bilinguals, but as L2 for Spanish-English bilinguals. Conversely, lexical forms in Spanish (bottom row of the figure) serve as L2 to one group but as L1 to the other. Bidirectional influences are therefore examined across the two groups of bilinguals. Although the figure captures the full set of options for the mapping between English and Spanish, we can not make direct comparisons across the two experiments. Therefore, although the comparison between the ‘split condition’ (rightmost column) and the ‘different-translation’ (leftmost column) in fact reflects the effect of a shared English translation, these comparisons are not available in the current study because the stimuli and paradigms are different in these two conditions. To illustrate, the Spanish pair ‘fruta’ and ‘plato’ correspond to different translations in English (‘fruit’ and ‘dish’, ‘different translation’ condition) and the Spanish pair ‘cita’ and ‘fecha’ correspond to a single ambiguous English word (‘date’, ‘split translation’ condition), but in the current study ‘different translation’ pairs are tested via two English words (‘fruit’ and ‘dish’), and the ‘split condition’ is tested via a single English word with two expressions instantiating its meanings (‘dinner date – expiration date’). These differences make it difficult to estimate the effect of a shared English translation in the current study.

Nonetheless, the current set of experiments attempts to provide a comprehensive picture of the pattern of cross-language influences that arise in the form-meaning mapping of bilingual speakers. We test a large set of words, embedded in sentence context or in phrasal context, which

vary greatly in their baseline semantic relatedness. We examine if the relatedness of these words changes as a function of translation status in a bilinguals' other language. Across two groups of bilinguals we attempt to identify cross-language influences in both the L1 to L2 direction and the reverse, and hope to shed more light on how these effects unfold over time.

## **2.0 EXPERIMENT 1: SHARED TRANSLATION EFFECT IN SENTENCES USING EYE TRACKING**

### **2.1 INTRODUCTION**

The mapping of words to meanings is often indirect such that a given word form corresponds to multiple referents. When these referents are related in meaning, the ambiguous word is said to be polysemous (e.g., line), and when the two referents are unrelated in meaning, the ambiguous word is said to be homonymous (e.g., ‘bark’). For bilingual speakers, this semantic ambiguity is amplified by language differences in the mappings of words to meanings, such that a given language may use two lexical forms to denote these different senses or meanings (e.g., in Spanish, ‘ladrido’ denotes the sound a dog makes and ‘corteza’ denotes the outer layer of a tree). These language differences result in translation ambiguity, such that two words in one language map onto a single shared translation in another. As reviewed in Chapter 1, other sources, such as synonymy and part-of-speech differences may also contribute to the prevalence of translation ambiguity. Of interest in the current investigation is whether translation ambiguity has consequences for bilingual meaning representations. That is, would the meanings of two words that share a translation in another language known to a bilingual speaker become more (or less) related by virtue of their shared lexical form in the other language. To illustrate, would bilingual speakers of English and Spanish consider ‘test’ and ‘proof’ to be more related in meaning

because the same word in Spanish ('prueba') can refer to both? We will refer to such an influence of a shared lexical form on the other language as the 'shared-translation effect'.

As will be reviewed below, the shared-translation effect is important in demonstrating semantic transfer between languages of bilingual speakers. If the mapping of form to meaning in one language changes the semantic relatedness of words in another language, then the bilingual mental lexicon is clearly interconnected. Moreover, the presence of bidirectional influences of a shared translation, from L1 to L2 and importantly from L2 to L1, provides a critical demonstration of the dynamic nature of the bilingual lexicon, one in which language experience continues to guide and shape organization and connections.

In the current experiment, we investigate the manifestation of the shared-translation effect in sentence context, by examining if bilingual speakers consider a shared-translation replacement to create less of a violation than a control different-translation replacement. To this end, English sentences were constructed to bias a particular target word, that is then replaced by a shared-translation or a different-translation counterpart. For example, in the sentence "The car accident was not my fault this time" the target word ('fault') is replaced with its shared-translation counterpart ('guilt', both translates to Spanish as 'culpa') or with its matched different-translation word ('anger'). As will be discussed below, both replacements should create anomalous sentences, but the degree of anomaly may vary for bilingual speakers.

In what follows we first review the evidence regarding the shared-translation effect. We then review studies in which a similar semantic anomaly manipulation was used to explain how reading behavior could reflect different levels of semantic anomaly. We conclude the introduction by outlining the predictions for the current experiment.

### **2.1.1 The shared-translation effect**

In two studies, Jiang (2002; 2004) examined the influence of shared translations in L1 on L2 processing. Specifically, Chinese-English bilinguals were asked to rate the semantic similarity of related English word pairs that either shared a translation in Mandarin (e.g., problem – question, both corresponding to ‘wenti’) or corresponded to different Mandarin translations. In comparison to monolingual English speakers, bilinguals rated shared-translation pairs as more similar in meaning than different-translation pairs. Additionally, in a timed relatedness judgment task, Chinese-English bilinguals were faster to indicate shared-translation pairs were related in meaning than different-translation pairs, despite comparable semantic relatedness for the two conditions based on monolingual English speakers. Furthermore, bilinguals more often misclassified a related English pair as unrelated when the two English words did not share a Chinese translation. These relatedness-judgment task results were replicated with Korean-English bilinguals (Jiang, 2004), who similarly showed an effect of L1 shared translations on relatedness judgments in L2. Jiang took these findings to support the ‘L1 Lemma Mediation’ hypothesis (Jiang, 2000), by which most L2 words access a copy of the L1 semantics, such that the meanings of words in L1 influence processing of L2 words (even of advanced learners). Critically, however, the possible influence of L2 shared-translation on semantic relatedness of L1 words was not explored in these studies.

In a recent study with Hebrew-English and English-Hebrew bilinguals, Degani, Prior, and Tokowicz (2011) observed higher relatedness ratings for English word pairs that share a Hebrew translation (e.g., home – house, which share the translation ‘bait’) compared to different-translation pairs. This shared-translation effect was demonstrated not only for Hebrew-English bilinguals who learned English as an L2, but critically also for English-Hebrew bilinguals who

learned Hebrew as an L2. Thus, for these bilinguals, a later-learned L2 (Hebrew) exerted an influence on semantic relatedness ratings of word pairs in the L1 (English). Further, this study extended the investigation beyond related word pairs (e.g., home-house; as in Jiang, 2002; 2004) to unrelated word pairs (e.g., tool-dish). The shared-translation effect was present, and of similar magnitude, for both types of items, suggesting that even relatively unrelated words (like ‘tool’ and ‘dish’) become more similar in meaning for bilingual speakers due to the shared Hebrew translation (‘kli’).

Increased relatedness for shared-translation words across different levels of baseline semantic relatedness and for both L1 and L2 words, might be due to co-activation in the interconnected bilingual lexicon. Specifically, Degani et al. (2011) proposed that when bilinguals encounter a shared-translation word (e.g., the Hebrew word ‘kli’ corresponding to both ‘tool’ and ‘dish’ in English), its two meanings receive activation (e.g., Elston-Güttler & Friederici, 2005; Onifer & Swinney, 1981). Likewise, its two translations are also likely to be activated (e.g., Schwartz & Fontes, 2008) because both languages of bilingual speakers tend to be activated even in a single language context (e.g., van Hell & Dijkstra, 2002). Based on Hebbian principles (Hebb, 1949), this co-activation of the two meanings and two translations of the shared-translation word will lead to increased connections between them. The semantic relatedness of the two words is therefore increased, giving rise to the shared-translation effect.

Alternatively, it can be argued that two words that share a translation might be *less* related because they tend to serve as lexical competitors in many contexts. For instance, when one talks about the wonderful weather in the spring, only the seasonal meaning of the word ‘spring’ is appropriate, and not the mechanical/coil meaning of the word (Chwilla & Kolk, 2003). By extension, only the translation corresponding to the appropriate meaning is relevant

(i.e., 'primavera' and not 'resorte'). The two meanings (and translations) are therefore mutually exclusive and may develop inhibitory connections. This account was proposed to explain reduced priming between two English words that map onto a shared homonym German translation. In particular, Elston-Güttler, Paulmann, and Kotz (2005b) examined whether processing of English words is influenced by the existence of a shared German homonym. They further examined if meaning dominance (i.e., how frequently each meaning is thought of in association to the ambiguous word) and the degree of ambiguity in English, as well as participants' proficiency, modulate the effects. In an all English task, participants were presented with target words (e.g., jaw) for lexical decision following either a shared-translation prime (e.g., pine-jaw, both corresponding to 'Kiefer' in German) or an unrelated prime (e.g., oak-jaw). They recorded both behavioral measures (RT and accuracy) and event-related-potentials (ERPs) while higher- and lower-proficiency German-English bilinguals performed the task. The results showed facilitation for the shared-translation prime in the accuracy data, but a tendency toward inhibition in the RT data. Specifically, higher proficiency German-English bilinguals showed longer RTs in the shared-translation condition compared to controls (i.e., reversed priming) when the dominant meaning served as the target. Interestingly, the ERP record revealed that lower-proficiency English-German bilinguals showed a less negative N200 for shared-translation primes relative to controls in the homonym condition, but a more negative N200 for shared-translation primes when the prime was ambiguous in English (e.g., duty-inch, both corresponding to 'zoll' in German). Thus only lower-proficiency bilinguals showed any modulation of the ERP record, and it was consistent with facilitation of the shared-translation prime for homonyms, but with inhibition of the shared-translation prime when it was also ambiguous in English. When the target was preceded by a sentence context (e.g., 'The beautiful

table was made of solid pine/oak... jaw'), the lower-proficiency bilinguals took longer to indicate lexicality for the target word when it was preceded by a sentence ending in a shared-translation word (i.e., reversed priming) compared to controls. Further, reversed priming was also observed in the N200, with a more negative N200 following shared-translation primes relative to controls. Higher-proficiency bilinguals did not show significant RT or N200 effects. The authors suggested that increased L2 proficiency allowed these bilinguals to exert better control over the activation of their languages, such that in the presence of a biasing semantic context in an all-L2 task there were no detectible influences of the L1. The lower-proficiency bilinguals, in contrast, exhibited what seems to be inhibition between the two meanings of a shared L1 homonym when these were presented in context. The authors suggested that lexical-level inhibitory connections develop between the two translations of homonyms, because these serve as lexical competitors in many contexts (e.g., Chwilla & Kolk, 2003).

Interestingly, the pattern of results obtained by Elston-Güttler et al., (2005b), namely inhibition between two translations of a shared translation, was not solely due to the presence of sentence context. In fact, in a different study examining the shared-translation effect in sentence context, Elston-Güttler and Williams (2008) found evidence consistent with increased relatedness of shared-translation words. In particular, advanced learners of English who were native speakers of German were asked to read English sentences, and decide whether the last word made sense as a completion of the sentence. For example, following the sentence 'His shoes were uncomfortable due to a', the word 'blister' was presented, and participants had to indicate as quickly and accurately as possible whether this last word was a good fit to the sentence. Critically, on some 'no' trials this last word shared a German polysemous translation with the intended word (e.g., 'bubble', both corresponding to 'Blasé' in German). In comparison

to native English speakers, learners showed a higher error rates and slower RTs to these shared-translation replacements. This suggests that shared-translation replacements were less anomalous for bilingual speakers. The authors further examined whether the effect was modulated by the degree of relatedness of the two senses of the ambiguous German word. Note that in all cases the two senses were semantically related (i.e., no German homonyms were included) but the degree of relatedness varied. They reasoned that if the effects were based on lexical level translational links between L1 and L2 words, the effects should be insensitive to the semantic relatedness of the two words. If, as in fact was observed for noun targets, learners showed stronger interference for highly-related senses, this would suggest at least partial activation of meaning. Based on their findings, the authors suggested that upon encountering an English word, its shared German translation is activated, which in turn activates its two senses. The activation of both senses interferes with bilinguals' ability to judge one of them as inappropriate to the sentence context.

This explanation assumes activation of the non-target language, but the possibility that this co-activation of words and meanings lead to changes in semantic representations was not considered. Further, the potential for L2 word-meaning mappings to influence processing of L1 words was not examined.

Thus, there is conflicting evidence with respect to the nature of influence of the shared-translation effect (inhibition or facilitation between shared-translation words) that is not reducible to the effect of sentence context. Increased relatedness or facilitation has been observed in isolation (Degani et al., 2011; Jiang, 2002; 2004) and in sentence context (Elston-Güttler & Williams, 2008), and when the shared-translation words were related in meaning (e.g., home-house, Degani et al., 2011; Elston-Güttler and Williams, 2008; Jiang, 2002; 2004) and unrelated in meaning (e.g., tool-dish, Degani et al., 2011). Inhibition has nonetheless been observed for

unrelated shared-translation words in sentence context (Elston-Güttler et al., 2005b). The results from the current experiment will address this issue and will potentially help reconcile these findings.

Results from several other studies are relevant to the shared-translation effect (Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011; Thierry & Wu, 2007; Wu & Thierry, 2010). In these studies, the influence of L1 on processing of L2 words was interpreted to reflect online unconscious and automatic activation of the non-target translation during the task, but the results could also be conceptualized as due to a (partial) shared-translation. Specifically, Thierry and Wu (2007) showed that Chinese-English bilinguals exhibited reduced N400 effects for pairs of English words for which the Chinese translations shared a character. These ‘partial-shared-translations’ elicited priming in the N400 window during a semantic relatedness judgment task. Interestingly, semantic relatedness was crossed with Chinese partial-shared-translation, such that related pairs (e.g., post-mail) and unrelated pairs (e.g., train-ham) could be connected via an overlap in a Chinese character. Across both levels of relatedness, bilinguals showed priming for the partial-shared-translation, and this N400 modulation was smaller and shorter than the semantic relatedness effect within the same time window. The effect of partial-shared-translation was evident in the ERP record during both visual and auditory semantic judgment tasks, but was not present in the RT or accuracy data from the same tasks. The authors interpreted their findings to suggest unconscious automatic activation of the Chinese translations. However, the N400 modulation is in fact also consistent with changes in the semantic representations of these words due to the partial-shared-translation. The lack of effect in the behavioral measures may be due to reduced power with only 15 participants.

In a second study, Wu and Thierry (2010) again examined the influence of Chinese (L1) translations on the processing of English (L2) word pairs, but now instead of crossing semantic relatedness and a shared Chinese character (as in Thierry & Wu, 2007), they examined the influence of Chinese orthographic overlap separately from phonological overlap and from semantic relatedness. Using the same semantic relatedness judgment task in the visual and the auditory modalities, the results showed an effect of Chinese phonological overlap on the N400 ERP component for Chinese-English bilinguals. English word pairs for which the Chinese translations overlapped in sound elicited a reduced N400 compared to those that overlapped in orthography and to unrelated controls, which did not differ from each other. As in Thierry and Wu (2007), no effects emerged in behavioral measures. These results were again interpreted in a processing framework, suggesting that bilinguals spontaneously activate the sound of L1 translations while processing L2 words. The possibility that shared phonological forms may have led to increased semantic relatedness for these items was not discussed. Moreover, neither study examined the possible influence of L2 on L1 processing.

Morford et al. (2011) similarly examined the influence of a shared form in L1 translations on L2 word processing. Interestingly, participants were bimodal bilinguals, who were deaf signers of ASL as L1 and had English as L2. They employed the semantic relatedness judgment task (Thierry & Wu, 2007; Wu & Thierry, 2010) and asked their participants to make timed relatedness judgments to pairs of English words that were related or unrelated, and for which the ASL translations shared or did not share phonological parameters (e.g., handshape, location, movement, and/or orientation). They found that deaf ASL-English bilinguals were faster to indicate related pairs were related when they shared phonology in ASL, and conversely were slower to indicate that unrelated pairs were unrelated when they shared phonology in ASL. This

pattern of results contrasted not only with that of monolingual English speakers (whose performance was used to select the critical stimuli) but also with that of hearing L2 English learners. This latter finding suggests that the effect was not due to general differences between bilinguals and monolinguals but rather due to the shared phonology in L1. Interestingly, the authors interpreted their findings to suggest activation of the non-target language (ASL) during the task, and like Thierry and Wu (2007; Wu & Thierry, 2010) did not discuss the possibility that the results reflect a stable change in semantic representations due to the shared phonology in ASL. Notably, stimuli were rated for semantic relatedness by five deaf ASL-English bilinguals, but these ratings served to ensure consistent allocation of the stimuli to the related versus unrelated conditions, and were not used to match the stimuli on semantic relatedness per se. Thus, it is possible that for bilingual speakers, English pairs that share phonology in ASL are more semantically related than pairs that do not share phonological parameters. Moreover, the pattern of error rates show that semantically unrelated pairs that share phonological parameters in ASL were often responded to as ‘related’ by the bilingual speakers. Rather than considering these as errors, one could interpret these trials as reflecting increased relatedness for the items for bilingual speakers, making a ‘yes’ response more appropriate. Such a pattern would be consistent with differences in semantic relatedness for bilinguals as a function of a partial-(phonological) shared-translation.

Finally, bilingual production errors provide suggestive evidence for the influence of a shared-translation on the meanings bilinguals assign to words. For instance, when a Spanish-English bilingual mistakenly says ‘thank you for your comprehension’ instead of ‘thank you for your understanding’ it is likely because a shared Spanish translation (‘comprensión’) corresponds to both ‘comprehension’ and ‘understanding’. Several investigations have shown

that bilinguals' production errors sometimes reflect a shared-translation effect. For example, as mentioned in Chapter 1, in a narrative analysis of oral productions of Russian-English bilinguals, Pavlenko and Jarvis (2002) observed 'semantic extensions', which occur when a bilingual uses the inappropriate translation of a shared-translation word. For example, bilinguals used the word 'neighbor' for 'roommate', because the Russian shared translation 'sosed' captures both. Marian and Kaushanskaya (2007) similarly identified bidirectional cross-language transfer in bilingual production data, when bilinguals extended the meaning of a word to encompass another word, based on the semantic content of the shared translation. These production data provide suggestive evidence that the shared-translation effect may exert its influence in both directions (L1 on L2 and L2 on L1).

The present experiment seeks to extend the reviewed findings in several ways. First, it is important to replicate the shared-translation effect across different language pairings to rule out the possibility that the effects are due to inherent differences between the items that share or do not share a translation across a certain language pair. Such replication is especially important because the influence of a later-learned L2 shared translation on L1 processing has been demonstrated only once (Degani et al., 2011). Second, because L2 on L1 effect was evident for bilinguals who had been immersed in their L2 for 20 years on average, it is critical to examine if a bidirectional pattern of the shared-translation effect depends on such immersion experience, or whether it can emerge even for bilinguals who are living in an L1 environment. The presence of bidirectional effects for such bilinguals will serve as a stronger test of the shared-translation effect. Third, the present experiment will examine the generalizability of the shared-translation effect across different levels of semantic relatedness in the same experiment, including a continuum ranging from unrelated pairs (corresponding to a homonym like 'vela' in Spanish,

meaning ‘candle’ and ‘sail’) to highly related pairs (corresponding to a polysemous Spanish word like ‘silla’, denoting both ‘seat’ and ‘chair’). We test these aspects of the shared-translation effect when the words are embedded in a sentence context by recording participants’ eye movements.

### **2.1.2 Semantic anomaly in eye tracking studies**

The current experiment takes advantage of a semantic anomaly paradigm, by which the semantic relatedness of shared-translation words can be estimated based on the degree of disruption created during reading when one word is replaced with its shared-translation (or different-translation) counterpart. We reason that if shared-translation words are more related in meaning due to their shared lexical form, they should be considered more interchangeable in a sentence context than controls. Alternatively, if the shared-translation words are less related, they should be considered less interchangeable and should create greater violation compared to controls. As mentioned above, Elston-Güttler and Williams (2008) used the same reasoning in their study, in which participants indicated if the last word in a sentence (either a shared-translation or a control) was a good fit to the sentence. Note, however, that in that study participants were engaged in a task that may have influenced their strategy during reading (for discussion of the influence of task on reading behavior, see also Warren, 2011).

In the current experiment we ask participants to read sentences, and record their eye movements as they do so. The eye tracking methodology employed here has the advantage that it does not require participants to perform any additional task. By manipulating whether the anomaly is created due to a shared-translation replacement or a control (different-translation)

replacement, we can estimate whether shared-translation words are more or less related in meaning, and more or less interchangeable in sentence context.

The eye movement record also provides a continuous record of how processing unfolds over time. In particular, several measures can be derived as participants read the sentences, reflecting earlier or later processing stages. *First-fixation duration* reflects the duration of the first fixation on a particular sentence-region during first-pass reading. *Gaze duration* reflects the sum of all fixations on a region during first-pass reading before the region is left for the first time. *Regression out* refers to the percentage of trials on which the reader leaves the region to the left (i.e., with a regressive saccade) during first-pass reading. *Skipping rate* refers to the percentage of trials on which the region was skipped during first-pass reading. These four measures are considered to reflect early processing. *Go-past time* refers to the cumulative duration of all fixations on a region from when the region is initially fixated during first-pass reading until the eyes leave the region to the right (i.e., with a progressive saccade). It is considered a measure of early processing by some (e.g., Murray, 2006), but as reflecting later stage processing by others (e.g., Libben & Titone, 2009; van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). *Total viewing time* refers to the cumulative duration of all fixations on the region, and combines both first- and second-pass reading, and reflects relatively late processing. In addition to these various measures, one can examine each of these measures on different regions of the sentence. The region at which the anomaly can first be detected is typically the *target region*, and the words following that point in the sentence are the *post-target region*, whereas those preceding it are the *pre-target region*.

Several studies have demonstrated the sensitivity of the eye-tracking method to disruptions due to semantic anomalies in sentence processing (e.g., Patson & Warren, 2010;

Rayner, Warren, Juhasz, & Liversedge, 2004; Warren & McConnell, 2007; Warren, McConnell, & Rayner, 2008; for review, see Warren, 2011). Moreover, these studies show that eye movement measures can reflect differences in the *degree* of semantic anomaly, which can be seen in both the onset and duration of the disruption (for review, see Warren, 2011).

For example, Rayner et al. (2004) examined readers' eye movements when reading sentences with semantic anomalies of different severities. At the less severe level, the anomaly reflected implausible events (e.g., 'John used an axe to chop the large carrots for dinner') and at the more severe level, the anomaly reflected events that are not only implausible, but are also extremely unlikely if not impossible (e.g., 'John used a pump to inflate the large carrots for dinner'). Relative to plausible control sentences, semantic anomalies increased go-past reading times on the word at which the anomaly became apparent (i.e., the target region) as well as on words following the anomalous word (i.e., post-target region). Moreover, the severity of the anomaly influenced both the timing at which the anomaly was detected and its duration. Specifically, the more severe condition disrupted reading as soon as the eyes landed on the target region (as reflected in longer gaze durations) and the effect lasted through the post-target region (as reflected by all measures). This study suggests that the eye tracking record is a sensitive measure not only of the presence of a semantic anomaly per se, but critically also of the degree of anomaly. This property is of great relevance in the current experiment, in which we seek to identify differences between what might be two levels of semantic anomalies (i.e., the shared- vs. the different-translation replacement conditions) for bilingual speakers.

Warren and McConnell (2007) similarly showed that the degree of semantic anomaly influenced the time at which the violation was detected and its duration. Participants read sentences that were possible and implausible (e.g., 'The man used a blow-dryer to dry the thin

spaghetti yesterday evening’) or reflected a stronger anomaly by being both impossible and implausible (e.g., ‘The man used a photo to blackmail the thin spaghetti yesterday evening’). In comparison to possible sentences, disruption in the severe impossibility anomaly condition was evident upon the first encounter with the target word (in first-fixation duration), and this disruption continued into the post-target region. The less severe implausibility anomaly, in contrast, emerged in slightly later measures on the target region, those that include regressions (i.e., regression out and go-past time), and these effects did not continue into the post-target region. It therefore appears that disruptions due to semantic mismatches are reflected in participants’ eye movements as they read sentences for comprehension.

### **2.1.3 The present experiment**

The present experiment focuses on the manifestation of the shared-translation effect in sentential context. Spanish-English (SE) and English-Spanish (ES) bilinguals read English sentences while their eye movements are recorded. Critically, we examine if a shared-translation replacement (e.g., the word ‘guilt’ instead of ‘fault’, which share the Spanish translation ‘culpa’) is more interchangeable with the target than a control different-translation word (e.g., ‘anger’ instead of ‘fault’). Similar to the sentence anomaly task employed by Elston-Güttler and Williams (2008), we expect bilinguals to show differences in anomaly processing when a shared-translation is used versus when a different-translation replacement is used. If shared-translation words become more similar in meaning (as is predicted by the co-activation account; Degani et al., 2011), then sentences in which a target word is replaced by a shared translation should lead to less difficulty in reading than sentences in which the same target word is replaced by a control different-translation word.

Alternatively, if two words that share a translation come to inhibit each other because they typically serve as lexical competitors (e.g., Elston-Güttler et al., 2005b), then a shared-translation replacement sentence should be more difficult to process than a different-translation replacement sentence. Under both accounts, both types of replacements are predicted to cause difficulty in comparison to the original sentence (in which the target word is not replaced). Critically, these condition effects (difference between shared-translation replacement and different-translation replacement) are predicted to vary as a function of language background group, such that differences between the two replacement types would be present only for bilingual speakers. Monolingual English (ME) speakers will provide a baseline to which bilingual performance can be compared.

Further, the direction of the condition effect (i.e., facilitation or inhibition for shared- vs. different-translation replacements) may change over time. In particular, comprehension processes unfold as participants read the sentence. Different patterns may emerge across different regions of the sentence (i.e., pre-target, target, and post-target regions). Moreover, the different measures derived from the eye movement record may elucidate different processes at different time points. For example, first-fixation durations on a region would index earlier processes than total-viewing time of the same region. We will therefore examine how the shared-translation effect unfolds over time within a region (across different eye movement measures) and across regions within a sentence.

Importantly, the current experiment further examined the degree to which the shared-translation effect is modulated by the baseline relatedness of the items. As mentioned above, two words that share a translation can vary in the relatedness of their meanings even for monolingual speakers. Some are unrelated (e.g., ‘doll’ and ‘wrist’, sharing the Spanish homonym ‘muñeca’),

some are moderately related (e.g., ‘proof’ and ‘test’, sharing the Spanish word ‘prueba’) and some are very highly related (e.g., ‘award’ and ‘prize’, for ‘premio’). In the current experiment, relatedness was operationalized as a continuum based on the ratings of monolingual English speakers. This allowed us to examine if the difference in processing shared- and different-translation replacements changes as a function of baseline semantic relatedness.

The present experiment further examines if the shared-translation effect is limited to the influence of L1 on L2 (e.g., Elston-Güttler et al., 2005b; Elston-Güttler & Williams, 2008; Jiang, 2002; 2004) or whether a bidirectional pattern may surface, such that a later-learned L2 would influence L1 processing (Degani et al., 2011). As mentioned above, two groups of bilinguals were examined. Both groups read English sentences and the effect of a shared-translation in Spanish was investigated. For SE bilinguals, a shared-translation effect will represent an influence of L1 on L2 during immersion in an L2 environment. This direction of influence was investigated in previous studies (e.g., Jiang, 2002; 2004), but was typically examined when bilinguals were immersed in their L1 (Degani et al., 2011; Elston-Güttler et al., 2005b). For ES bilinguals, a shared-translation effect will reflect the influence of a later learned L2 (Spanish) on reading behavior in the L1 (English), for bilinguals who are currently immersed in their L1. Such L2 on L1 influence was investigated only once (Degani et al., 2011), and in that study bilinguals were immersed in their L2. Results from both groups will therefore shed light on the relevance of immersion environment on the manifestation of the shared-translation effect.

Because previous research has clearly demonstrated the importance of lexical factors (e.g., length, frequency) and semantic constraints (i.e., predictability) on fixation durations and probability of skipping (for a recent demonstration of the independent contribution of length and predictability, see Rayner, Slattery, Drieghe, & Liversedge, 2011), it was imperative to match

these dimensions across the shared-translation and different-translation replacements. Note that length was not necessarily matched to the intended target, but importantly was matched between the shared-translation and different-translation replacements on an item by item basis. Furthermore, careful norming experiments with monolingual English speakers were conducted to identify any potential differences in word predictability and naturalness between the two critical anomalous conditions. These were later partialled out in a statistical model, such that any differences between the shared-translation replacement and the different-translation replacement for bilinguals compared to monolinguals should be due to the translation status of the words in Spanish, and not to other potentially confounding factors.

To summarize, the goals of the present experiment are to examine (1) the generalizability of the shared-translation effect to sentential context; (2) the time-course of the effect as reflected in different sentence regions and different measures of the eye-movement record; (3) the manifestation of the shared-translation effect across a continuum of semantic relatedness; and importantly (4) the bidirectional pattern of the effect, examining if a later-learned L2 can influence semantic processing of L1 words for bilinguals who are immersed in their L1 environment.

## **2.2 METHOD**

### **2.2.1 Participants**

Thirty monolingual English speakers, 30 English-Spanish bilinguals, and 30 Spanish-English bilinguals took part in the eye-tracking experiment. See Table 1 for background information on

the final set of 90 participants. The monolingual English speakers (ME; mean age 19.83 years,  $SD=2.69$ ; 19 females) participated for class credit. All were native English speakers who were not exposed to other languages before age 10 (with the exception of one who reported having some exposure at age 7), and were exposed to English at least 80% of the time at the time of testing. None had learned Spanish.

The English-Spanish (ES) bilinguals were paid for their participation. All were native English speakers who had studied Spanish as an L2 and were not exposed to Spanish at home during childhood. They were at least moderately proficient in Spanish. The Spanish-English (SE) bilinguals were also paid for their participation. All were native Spanish speakers who learned English as an L2 and were not exposed to English at home during childhood. They were at least moderately proficient in English. Two participants had indicated that English had become their dominant language.

An additional 22 participants were replaced because they did not meet language background criteria (12 participants – 3 ME, 4 ES, 1 SE), achieved less than 75% accuracy on the filler comprehension questions in the eye tracking task (9 participants - 3 ME, 3 SE, 3 ES, see also Libben and Titone, 2009, for a similar criterion), or had difficulty following instructions on multiple tasks (1 SE participant). Calibration of the eye tracker failed for 7 additional participants (3 ME, 3 ES, 1 SE), and data from 3 other participants were not used to maintain counterbalancing across group and version (1 ES, 2 ME). One participant (ES bilingual) indicated having a reading disability, but because his accuracy on the filler reading comprehension questions was not extreme, his data were maintained.

**Table 1.** Background characteristics for the final set of participants by group

Measure	Linguistic Background Group		
	Spanish-English Bilinguals	English-Spanish Bilinguals	English Monolinguals
Number of participants	30 (8 males)	30 (4 males)	30 (11 males)
L1	Spanish	English	English
Age (years)	33.63 (11.84)a	23.83 (8.35)b	19.83 (2.69)b
Age began L2 (years)	10.98 (6.13)a	12.10 (3.09)a	n/a
Time studied L2 (years)	13.20 (7.44)a	9.15 (5.03)b	n/a
L2 immersion (years)	6.98 (8.44)a	0.33 (0.58)b	n/a
L1 proficiency	9.75 (0.53)a	9.91 (0.23)a	9.77 (0.48)a
L2 proficiency	8.06 (1.36)a	7.79 (0.79)a	n/a
L1 current use	5.19 (2.12)a	8.61 (1.33)b	9.75 (0.63)c
L2 current use	7.89 (1.30)a	4.03 (1.68)b	n/a
Attitude toward reading	9.30 (1.58)a	8.73 (1.68)ab	8.23 (1.79)b
Rated reading amount	7.54 (1.79)a	7.18 (1.56)ab	6.35 (1.24)b

*Note.* Proficiency scores are the average of reading, writing, conversational, and speech comprehension ability ratings on a 10-point scale, on which 1 indicated the lowest level of ability. Current use scores are the average of speaking, writing, reading, listening to the radio, and watching TV ratings on a 10-point scale on which 1 indicated the lowest level of current use. Attitude toward reading reflect ratings on a 10-point scale, on which 1 reflects a ‘very negative’ attitude, and 10 reflects a ‘very positive’ attitude. Reading amount scores reflect the average reading for pleasure, work, and school on a 10-point scale, on which 1 reflects ‘none’ and 10 reflects ‘a great deal’. Means in the same row that do not share sub-scripts differ at the  $p < .05$  level in a  $t$ -test with the Bonferroni correction for multiple comparisons.

### 2.2.2 Materials

One hundred and twenty critical English sentences were constructed to serve as sentence contexts to 120 word triplets. Each triplet was selected to include an *intended shared-translation* (ITS) word, a *shared-translation replacement* (STR) which shares a Spanish translation with the ITS, and a matched *different-translation replacement* (DTR). For instance, if the word ‘fault’ serves as the ITS, the word ‘guilt’, which shares with it the Spanish translation ‘culpa’, would serve as the STR, and the word ‘anger’ would serve as the DTR. For each word triplet a sentence frame was developed, such that it would be a natural-sounding sentence with the ITS completion but similarly unnatural-sounding with either an STR or a DTR completion. For example, in the sentence ‘The car accident was not my fault this time’ replacing the ITS ‘fault’ with either the STR ‘guilt’ or the DTR ‘anger’ creates similarly anomalous sentences. Additional examples are presented in Table 2. The full set of experimental sentences is available in Appendix C.

**Table 2.** Experiment 1 example sentences.

Sentence Frame	Target (IST)	Shared (STR)	Control (DTR)
The ticking hands of the grandfather ___ were the only sound in the house.	clock	watch	sleep
I'm sorry my comment offended you, but it was never my ____ to hurt you.	intention	attempt	message
To mask the smell of the burnt food, Alison lit a scented ____ in the kitchen.	candle	sail	jail

Stimulus selection was informed by an extensive set of norms collected from monolingual English speakers and bilinguals of Spanish and English (see Appendix A). In particular, word pairs were first normed for semantic and form similarity by monolingual English speakers, and were then normed for translation overlap by bilinguals of English and Spanish.

Based on these, a set of 120 word triplets that varied in their semantic relatedness were chosen. Experimental sentences were then developed, and selected based on an additional set of norms, which included sentence predictability norms and ratings of sentence naturalness, collected from monolingual English speakers. Stimuli characteristics are summarized in Table 3.

**Table 3.** Experiment 1 stimuli characteristics.

Measure	Target (IST)	Shared (STR)	Control (DTR)
Word Length (in letters)	5.88 (2.10)a	5.35 (1.62)b	5.38 (1.64)b
Word Log KF Frequency	1.71 (0.54)a	1.63 (0.54)a	1.61 (0.56)a
Word Concreteness Ratings	464.98 (119.94)a	451.09 (117.81)a	446.69 (116.62)a
Form Similarity to the IST	N/A	1.98 (0.93)a	2.06 (0.67)a
Semantic Similarity to the IST	N/A	3.98 (1.76)a	3.37 (1.28)b
Sentence Length (in characters)	68.70 (9.39)a	68.16 (9.17)a	68.18 (9.18)a
Sentence Predictability	5.59 (3.72)a	0.10 (0.35)b	0.00 (0.00)b
Sentence Naturalness Rating	2.15 (0.74)a	4.00 (1.20)b	4.72 (1.16)c

*Note:* Word log Kucera-Francis frequency and concreteness ratings were taken from the MRC database (Wilson, 1988). Form similarity scores are the average ratings on a 1-7 scale of the word and the IST (on which 1 reflects low similarity and 7 reflects high similarity). Semantic similarity score are the average ratings of the word and the IST on a 1-7 scale, on which 1 indicates ‘completely different’ and 7 indicates ‘exactly the same’. Sentence predictability scores reflect how often the word was provided as a completion to the sentence in the predictability norms (range of 0 to 10), and sentence naturalness ratings range on a scale of 1-7, where 1 indicates high naturalness and 7 indicates low naturalness. Ratings were collected from monolingual English speakers (see Appendix A). Means in the same row that do not share subscripts differ at the  $p < .05$  level in a  $t$ -test with the Bonferroni correction for multiple comparisons.

Experimental sentences were presented one at a time. Three versions were created, each with the same 40 filler sentences (taken from Patson & Warren, 2010). Each of the filler sentences was followed by a yes/no comprehension question. Experimental sentences were counterbalanced across the three versions, such that in each version a third of the sentences included the IST completion, a third included the STR completion, and a third included the DTR completion. At least half of the sentences in each version (IST completions and fillers) were therefore relatively natural sentences. Further, to allow variability in semantic relatedness, three levels of relatedness were established based on the paired semantic relatedness norms (see Appendix A). Approximately a third of the sentences in each translation condition were from the highly-related level, a third were from the moderately-related level, and a third were from the unrelated level. Sentence frames and target words were not repeated within the same version.

### **2.2.3 Procedure**

#### **2.2.3.1 General procedure**

Participants were informed that the study was comprised of multiple tasks, the first of which involved recording their eye movements as they read sentences in English. Bilinguals were told that some of the tasks following the eye tracking section would be in Spanish and some would be in English. This was done to reduce the chances that participants would actively search for the relevance of Spanish to the reading task, because they knew their knowledge of Spanish was relevant to their participation. Upon completion of the *sentence reading task*, participants performed a yes/no *semantic relatedness task* (see Experiment 2 – Chapter 3). They then completed a *naturalness rating* task of the sentences they had seen, to provide an off-line measure of the perceived sentence anomaly as a function of condition and relatedness.

Bilinguals then completed two proficiency tasks – a *lexical decision* task and a *picture naming* task (the latter taken from Tokowicz, 1997) in their L2, followed by the *Raven's Progressive Matrices* task (Ravens, 1960; see e.g., Landi, 2010; Nelson, 2010, for abbreviated version), which also served as a buffer task, before they performed the same proficiency tasks (lexical decision and picture naming) in their L1. Participants then completed the *operation-word working memory span* (O-Span) task in their L1 (Turner & Engle, 1989; see also Michael et al., 2011; Tokowicz, Michael, & Kroll, 2004), and filled out the *language history questionnaire*. Before debriefing, bilinguals also completed a *vocabulary post test*, to verify their familiarity with the critical word triplets.

The order of tasks for the English monolinguals was slightly different. Following the sentence naturalness task, they completed the lexical decision, picture naming, and O-span task in English. They then filled out the language history questionnaire and the Raven's Matrices task. Due to time constraints, some monolingual speakers did not complete the Raven's task. They were then debriefed about the study.

Because of the length of the study (2.5-3.5 hours for bilinguals; 1.5-2.5 hours for monolinguals), participants were encouraged to take breaks between tasks. The procedures and details of the sentence reading and sentence naturalness tasks are described below. Details on the procedures, materials, and results of the other tasks are available in Appendix D.

### **2.2.3.2 Sentence reading.**

Participants were asked to read the sentences silently, as they would naturally, while their eye movements were recorded. An EyeLink 1000 eye-tracker (SR Research Ltd) was used to monitor participants' right eye, with a spatial resolution of 0.01° and gaze location sampling at 1000 Hz. E-Builder software (SR Research Ltd.) was used for stimulus presentation. A standard 9 point

full-screen calibration was used at the beginning of the session, and recalibration was performed when needed. A 3 point drift correction procedure was used before each sentence.

Sentences were presented one at a time, vertically centered with a left justification. Each sentence was presented in one line (with no more than 80 characters per line) in black Courier New 14 point font on a white background, and participants were instructed to press a button when they were done reading the sentence. Forced choice comprehension questions followed one quarter of the sentences (filler natural sentences, see, e.g., Libben & Titone, 2009) to ensure that participants were reading the sentences for meaning. ‘Yes’ responses were made with the right hand and ‘no’ responses were made with the left hand. The eye tracking session lasted between 20 and 60 minutes.

### **2.2.3.3 Sentence naturalness rating**

Sentences were rated for naturalness on a scale of 1 (very natural) to 7 (very unnatural) via a web-based interface; the scale was reversed prior to analyses to allow more intuitive understanding of the findings. Participants were encouraged to use the full range of the scale, and were instructed to rate the sentences in the order in which they were presented. No time limit was imposed, and participants were informed that we were interested in their intuitions and that this was not a test of their knowledge. Further, they were told that the sentences would overlap with the ones they had seen during the eye-tracking session. There were three versions of this task, and version assignment was in alignment with that of the sentence reading task. Each version included only the critical sentences (120) from that version of the sentence reading task, presented in a randomized order.

## 2.3 RESULTS

### 2.3.1 Sentence reading task

#### 2.3.1.1 Data pre-processing and trimming

*Accuracy on filler comprehension questions.* Performance on the comprehension questions was relatively high ( $M=.86$ ,  $SD=.05$ ). Note that participants were excluded if their accuracy was below .75 (9 participants: 3 ME, 3 SE, 3 ES; see also Libben & Titone, 2009, for a similar criterion). Analyses of Variance revealed a marginal difference among the groups in their comprehension accuracy,  $F(2,89)=2.48$ ,  $MSE=.002$ ,  $p=.089$ . Bonferroni corrections show that the SE group was marginally less accurate ( $M=.84$ ) than the ES group ( $M=.87$ ). The ME group did not differ from either other group ( $M=.86$ ).

*Eye movement data.* In each sentence, three critical interest areas were defined for analyses. The target region included the IST, STR, or DTR word within each sentence. The pre-target region included the word preceding the target, or in cases in which that word was less than 5 characters in length, it included the two words preceding the target. Similarly, the post-target region included the word following the target if it was at least 5 characters long, or else included the two words following the target word.

Eye tracking data were first cleaned to remove any fixations that were shorter than 80 ms or were longer than 1000 ms. Fixations that were outside the interest areas were moved vertically to correct for eye tracking drift. Fixations at the beginning of the trial that were not followed by a progressive saccade were deleted to maintain first-pass measures. Trials including blinks inside the pre-target, target, or post-target regions (during first-pass reading) were removed from the data. Blinks inside the interest areas led to the removal of 7.99% of trials. Analyses of Variance

showed that there were no significant differences in the proportion of trials removed due to blinks as a function of group,  $F < 1$ , or condition,  $F(2,174)=1.151$ ,  $MSE=17.451$ ,  $p=.319$ , with no interaction,  $F < 1$ .

The analyzed regions were fixated during first-pass reading on approximately 87.77% ( $SD=.75$ ) of the trials. These fixations varied, however, by condition and by group. In particular, IST sentences were fixated during first-pass reading less often ( $M=86.81$ ) than STR ( $M=88.18$ ) and DTR ( $M=88.32$ ) sentences,  $F(2,74)=7.75$ ,  $MSE=8.00$ ,  $p=.001$ . In addition, the ES participants fixated during first-pass reading less often ( $M=82.54$ ) than the ME ( $M=89.32$ ) and the SE participants ( $M=91.44$ ),  $F(2,87)=12.70$ ,  $MSE=153.20$ ,  $p < .000$ . Condition and group also interacted,  $F(4,174)=3.65$ ,  $MSE=7.997$ ,  $p=.007$ . Examination of the means revealed that whereas ME and ES participants fixated IST sentences less often than STR and DTR sentences during first-pass reading, SE bilinguals showed equivalent fixations across conditions. These differences across group and condition were examined more closely in the analyses of skipping probability reported below.

Approximately 9.1% of the data were removed prior to analyses due to participants' unfamiliarity with the shared-translation pair (i.e., the IST, STR, or both). This was done for bilingual participants only on an item-by-item basis, based on each participant's vocabulary post-test (see Appendix J).

### **2.3.1.2 Dependent measures**

Six measures were computed based on the cleaned eye tracking record for each of the three interest areas. As reviewed in the introduction, these include *First Fixation Duration (FFD)*, *Gaze Duration (GD)* (or first-pass time), *Go-Past Time (GPT)* (or regression path duration),

*Total Time (TT)*, *Regressions Out (RO)*, and *Skipping Rate (SR)*. Increased difficulty is generally associated with longer durations, more regressions out, and less skipping.

### **2.3.1.3 Analysis approach**

Analyses were performed on data from the final set of 90 participants using linear mixed effects models as implemented in the lme4 library in R (Baayen, Davidson, & Bates, 2008; R Development Core Team). The models included random effects for participants and items, and fixed effects for group (ME, ES, SE), translation (STR and DTR), relatedness (taken from the semantic similarity norms estimating the relatedness of the target word, STR or DTR, to the IST), and all the two-way and three-way interactions between them. In addition, other item and participant covariates were included in the models: target length, target log Kucera-Francis frequency, target part of speech (POS), log transformation of the form similarity of the target word (STR or DTR) to the IST, sentence predictability and sentence naturalness, as well as participant's age and mean performance on IST sentences (as described below).

Because we were interested in how bilinguals differ from monolinguals, for the group effect, the ME group were treated as the reference group such that we examined the difference between the ES group and ME group, and between the SE group and ME group. Because the difference between the two anomalous conditions is most relevant to the present experiment, for the effect of condition, we directly compared performance on STR sentences to DTR sentences. Note that sentences with IST completions were not included in the models. Instead, to control for potential baseline individual differences in reading behavior on these normal sentences, mean performance on IST sentences was computed for each individual in each measure on each sentence region, and these means were entered into the models as a covariate. The difference between STR and DTR sentences was therefore computed after taking into account reading

patterns on normal sentences for each individual. Note, also, that because IST sentences were not included in the models, the effects of the different predictors were estimated based on somewhat unnatural sentences (DTR and STR only), and should be interpreted as such.

Prior to analysis, relatedness and all covariates (with the exception of target part of speech) were centered. Dependent measures reflecting durations (i.e., first fixation duration, gaze duration, go-past time, and total time) were log transformed to reduce skewness in the distribution. For duration measures, significance was estimated based on 10,000 Markov Chain Monte Carlo samples of the posterior samples of the parameters (pMCMC, Baayen, 2008). Logistic models were fitted for the binomially distributed regression out and skipping data (Jaeger, 2008), yielding  $p$  values for significance estimation.

For each measure in each region, an additive model was fit first, followed by a model including the two-way interactions between group, condition, and relatedness. Finally, a model including the three-way interaction among these factors was tested. The reported coefficients are taken from the highest-level model that was significant or marginally significant. In what follows we report significant fixed effects for the theoretically-relevant predictors of group, condition, and relatedness for the target and post-target regions only. The full models are presented in Table 4 and Table 7. Detailed analyses of the pre-target region are available in Appendix E.

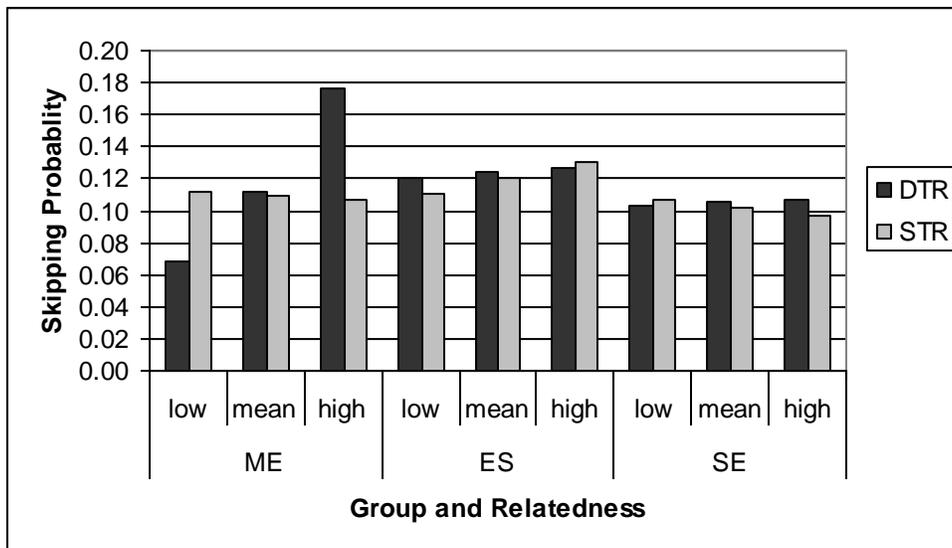
### 2.3.1.4 Target region

**Table 4.** Coefficient estimates for the target region

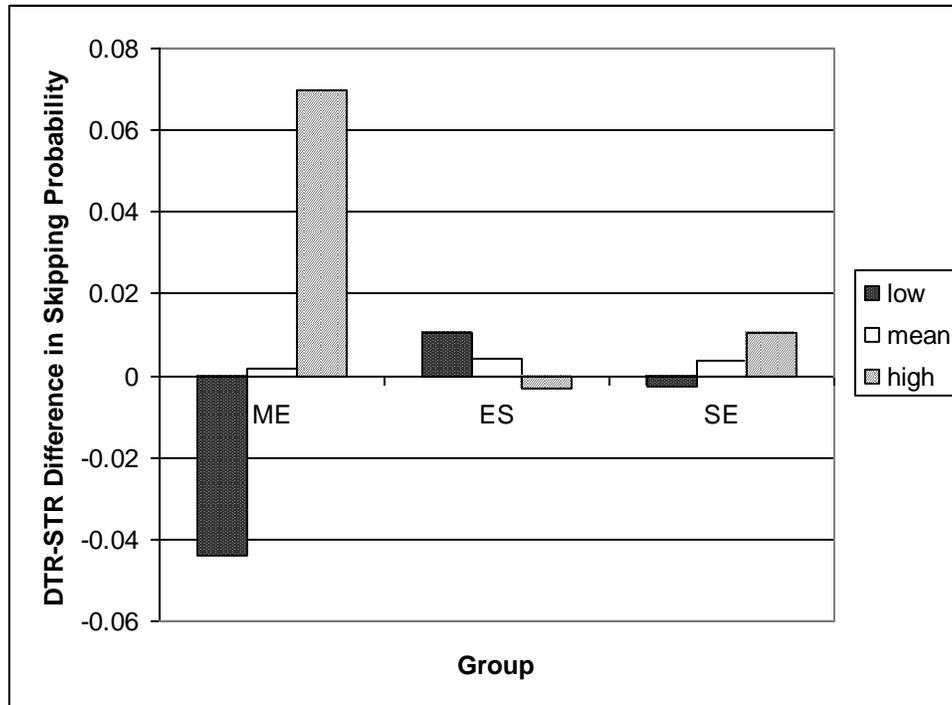
	SK	FFD	GD	GPT	TT	RO
Intercept	-2.076**	5.409**	5.512**	5.723**	5.929**	-1.495**
Participant Age	0.007	0.001	-0.001	0.002	-0.002	0.014*
Mean Performance on IST sentences	5.669**	0.004**	0.002**	0.001**	0.002**	4.512**
Target Log KF frequency	0.017	-0.003	-0.020**	-0.017*	-0.017*	0.019
Target Length (in letters)	-0.421**	0.001	0.026**	0.021*	0.052**	-0.065
Target Part of Speech [verb]	-0.017	-0.012	0.015	-0.056	0.029	-0.561*
Sentence Predictability Rating	0.173	-0.003	0.003	-0.014	0.025	-0.259
Sentence Un-naturalness Rating	0.003	0.009	0.014±	0.035**	0.072**	0.131*
Log Form-Similarity (to the IST)	0.197	-0.007	-0.018	-0.006	0.017	0.013
Semantic Relatedness (to the IST)	0.351**	-0.011	-0.022*	-0.034*	-0.049**	-0.055
Group [ES]	0.121	0.010	0.014	0.017	0.073	0.054
Group [SE]	-0.062	-0.001	0.036	-0.024	0.085	-0.719**
Condition [STR]	-0.019	-0.001	0.023	-0.017	-0.013	-0.384*
Condition [STR]:Group [ES]	-0.018	0.008	-0.014	0.001	-0.017	0.285
Condition [STR]:Group [SE]	-0.024	-0.014	-0.030	0.017	-0.035	0.529*
Condition [STR]:Semantic Relatedness	-0.369**	0.003	0.009	0.020±	0.025*	0.039
Group [ES]: Semantic Relatedness	-0.331*	0.005	0.005	0.002	-0.007	0.042
Group [SE]: Semantic Relatedness	-0.338*	0.015*	0.017*	0.024*	-0.009	0.060
Condition [STR]: Group [ES]: Semantic Relatedness	0.411*	-0.010	0.017	0.010	0.012	-0.071
Condition [STR]: Group [SE]: Semantic Relatedness	0.322*	-0.009	0.027	0.014	0.012	-0.131

Note: ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

*Skipping*: Overall, skipping of the target region significantly increased as the semantic relatedness of the target to the IST increased. This main effect and significant two-way interactions between condition and relatedness, and between relatedness and group, were qualified by a three-way interaction among condition, relatedness, and group. As can be seen in Figure 2, ME speakers tended to skip the target region more often in STR sentences when semantic relatedness was low, but more often in DTR sentences when semantic relatedness was high. By contrast, the ES bilinguals showed a reversed pattern, with more skips in DTR sentences when semantic relatedness was low and more skips in STR sentences when semantic relatedness was high. The SE bilinguals also differed significantly from the ME speakers, mostly in the magnitude of the difference between DTR and STR sentences across the range of semantic relatedness (see Figure 3). We address the condition differences within the ME speaker group in the Discussion section.



**Figure 2.** Skipping probability in the target region as a function of group, condition, and semantic relatedness.



**Figure 3.** Differences in skipping probability in the target region between DTR and STR sentences as a function of group and semantic relatedness.

*First-Fixation Durations:* The effect of relatedness on FFD varied by group, such that increased relatedness led to stronger reductions in FFD for SE bilinguals than for ME speakers (see Table 5).

*Gaze Durations:* Gaze duration in the target region significantly increased as the semantic relatedness of the target to the IST decreased. Further, as in the FFD data, the effect of relatedness varied by group (for both STR and DTR sentences) such that it was stronger for SE bilinguals compared to ME speakers (see Table 5).

*Go-Past Time:* GPT on the target region increased as the semantic relatedness of the target to the IST decreased. The effect of relatedness marginally interacted with condition, such that increased semantic relatedness decreased GPT slightly more in DTR sentences than in STR

sentences. Finally, the effect of relatedness varied by group, but here it was weaker for SE bilinguals, such their GPT decreased less with increased relatedness compared to ME speakers (see Table 5).

**Table 5.** Model estimates for FFD, GD, and GPT in the target region as a function of group and semantic relatedness.

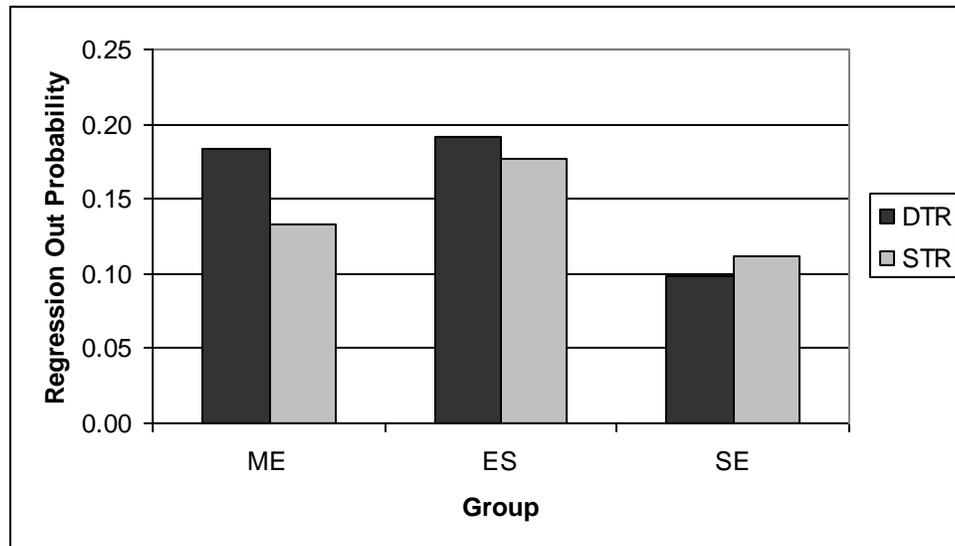
	Semantic Relatedness	ME	ES	SE
FFD	Low	227.05	226.40	231.63
	Mean	223.41	225.54	223.21
	High	219.83	224.68	215.09
GD	Low	256.10	265.19	277.94
	Mean	247.62	251.04	256.65
	High	239.42	237.64	236.99
GPT	Low	322.16	327.17	306.60
	Mean	305.94	311.22	298.72
	High	290.54	296.05	291.03

*Total Time:* TT on the target word significantly increased as the semantic relatedness of the target to the IST decreased, but the effect of relatedness again interacted with that of condition, such that it was stronger in DTR sentences than in STR sentences (see Table 6).

**Table 6.** Model estimates for TT in the target region as a function of condition and semantic relatedness.

Semantic Relatedness	DTR	STR
Low	389.80	384.92
Mean	375.85	371.15
High	362.41	357.87

*Regressions Out:* Overall, SE bilinguals regressed out of the target region less often than ME speakers, and DTR sentences elicited more RO than STR sentences, but critically these two factors significantly interacted. As can be seen in Figure 4, whereas ME speakers tended to regress out of the target region more often in DTR sentences, SE bilinguals tended to regress out more often in STR sentences.



**Figure 4.** Regression-out probability from the target region as a function of group and condition

*Summary of results in the target region:* Two measures indicate that bilinguals differ from monolingual speakers in their processing of STR and DTR sentences. Specifically, skipping of the target region varied as a function of group and relatedness, such that in contrast to ME speakers, ES bilinguals skipped the target region more often in DTR sentences when relatedness was low, but more often in STR sentences when relatedness was high. This suggests that bilinguals find low-relatedness STR sentences to be more difficult to process, but high-relatedness STR sentences easier to process than DTR controls. This finding is consistent with inhibition between unrelated words that share a translation (Elston-Güttler et al, 2005b) as well

as increased relatedness (possibly due to co-activation, Degani et al., 2011) between highly related words that share a Spanish translation. SE bilinguals show some divergence from ME speakers in the skipping data, but more prominently exhibit more regressions out of the target region on STR sentences, suggesting increased difficulty for STR sentences in comparison to DTR sentences. This finding was not modulated by degree of relatedness, and is consistent with inhibition or interference between two words that share a Spanish translation.

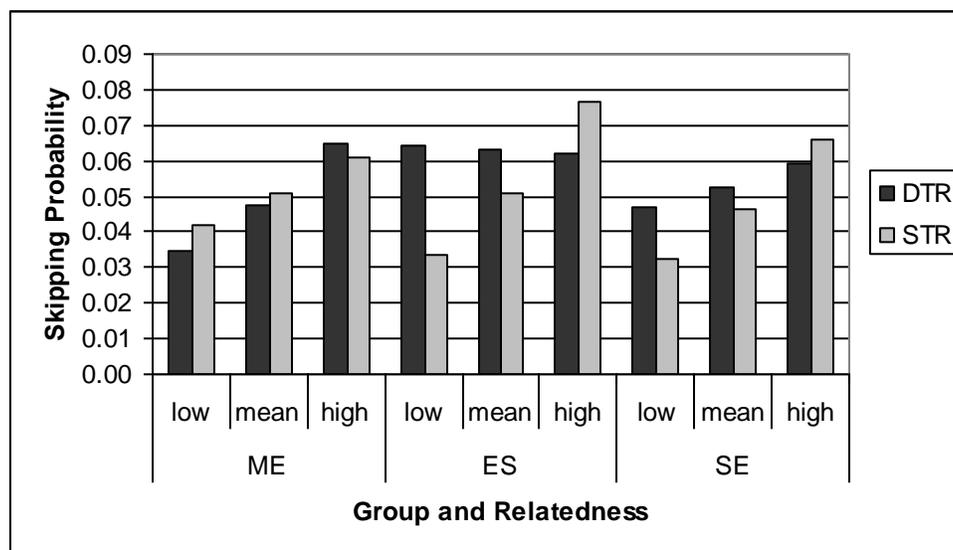
### 2.3.1.5 Post-target region

**Table 7.** Coefficient estimates for the post-target region.

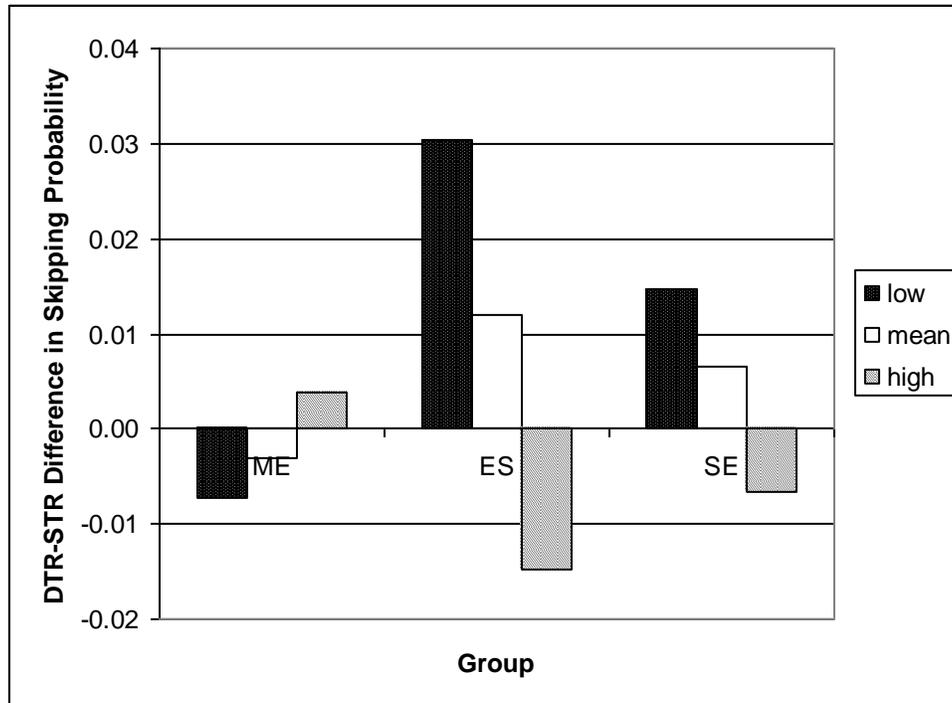
	SK	FFD	GD	GPT	TT	RO
Intercept	-2.999**	5.463**	5.816**	6.294**	6.304**	-0.815**
Participant Age	-0.005	0.001	0.001	0.006*	0.003	0.012
Mean Performance on IST sentences	7.031**	0.003**	0.002**	0.001**	0.001**	3.592**
Target Log KF frequency	-0.035	0.007	-0.002	-0.004	-0.005	-0.024
Target Length (in letters)	0.028	0.016*	0.019±	0.011	0.004	0.015
Target Part of Speech [verb]	-0.206	0.003	0.046	0.008	0.132*	0.031
Sentence Predictability Rating	0.400	0.049*	0.101*	0.051	0.044	0.068
Sentence Un-naturalness Rating	0.120±	0.006	0.015	0.054**	0.043**	0.247**
Log Form-Similarity (to the IST)	-0.107	-0.008	-0.010	0.058	0.048	0.237±
Semantic Relatedness (to the IST)	0.213±	-0.009±	-0.015	-0.012	0.007	-0.015
Group [ES]	0.300	-0.031	-0.083*	-0.135*	-0.105*	-0.120
Group [SE]	0.112	-0.043	0.001	-0.143*	-0.116*	-0.610*
Condition [STR]	0.067	-0.002	0.007	-0.032	-0.039	-0.003
Condition [STR]:Group [ES]	-0.290	0.000	0.016	-0.016	0.028	-0.166
Condition [STR]:Group [SE]	-0.205	-0.011	-0.005	-0.017	0.020	-0.201
Condition [STR]:Semantic Relatedness	-0.085	-0.008	-0.014	-0.017	-0.040*	0.018
Group [ES]: Semantic Relatedness	-0.225±	0.004	0.009	-0.048±	-0.022	-0.202*
Group [SE]: Semantic Relatedness	-0.132	0.001	0.005	-0.031	-0.039*	-0.145
Condition [STR]: Group [ES]: Semantic Relatedness	0.381*	-0.011	0.001	0.057±	0.049*	0.217±
Condition [STR]: Group [SE]: Semantic Relatedness	0.249	0.010	-0.005	0.029	0.040±	0.029

Note: ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

*Skipping*: Skipping rate in the post-target region marginally increased as the semantic relatedness of the target word to the IST increased. Of interest, this effect also marginally interacted with group, and in a three-way interaction with group and condition, such that when relatedness was low, ES bilinguals skipped the post-target region more often in DTR sentences than in STR sentences, but when relatedness was high, they skipped the post-target region more often in STR sentences than in DTR sentences. The ME speakers showed a reversed pattern, with smaller differences between conditions (see Figure 5). As is made clear in Figure 6, SE bilinguals patterned similarly to ES bilinguals, but the difference between them and ME speakers did not reach significance.



**Figure 5.** Skipping probability in the post-target region as a function of group, condition, and semantic relatedness.



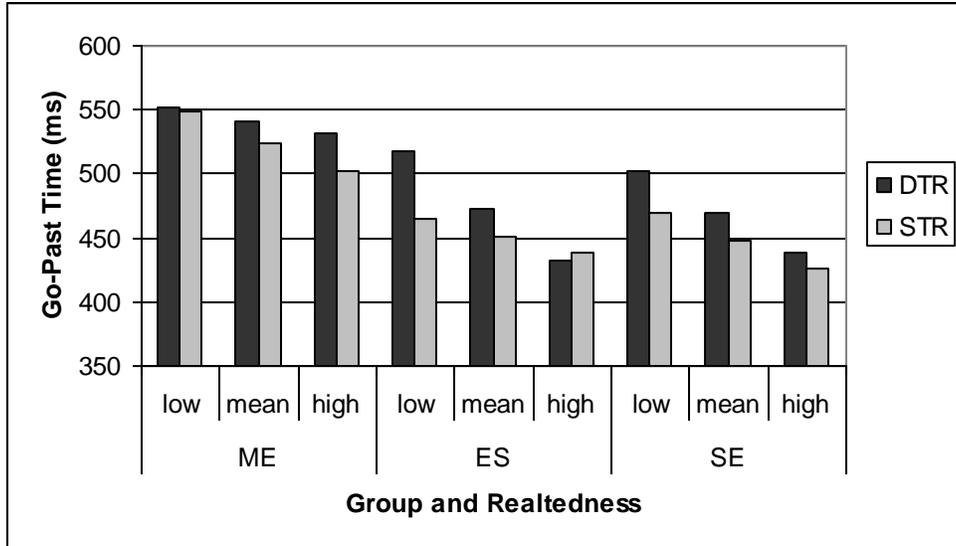
**Figure 6.** Differences in skipping probability in the post-target region between DTR and STR sentences as a function of group and semantic relatedness.

*First-Fixation Durations:* First fixation durations on the post-target region marginally increased as the semantic relatedness of the target word to the intended target word decreased.

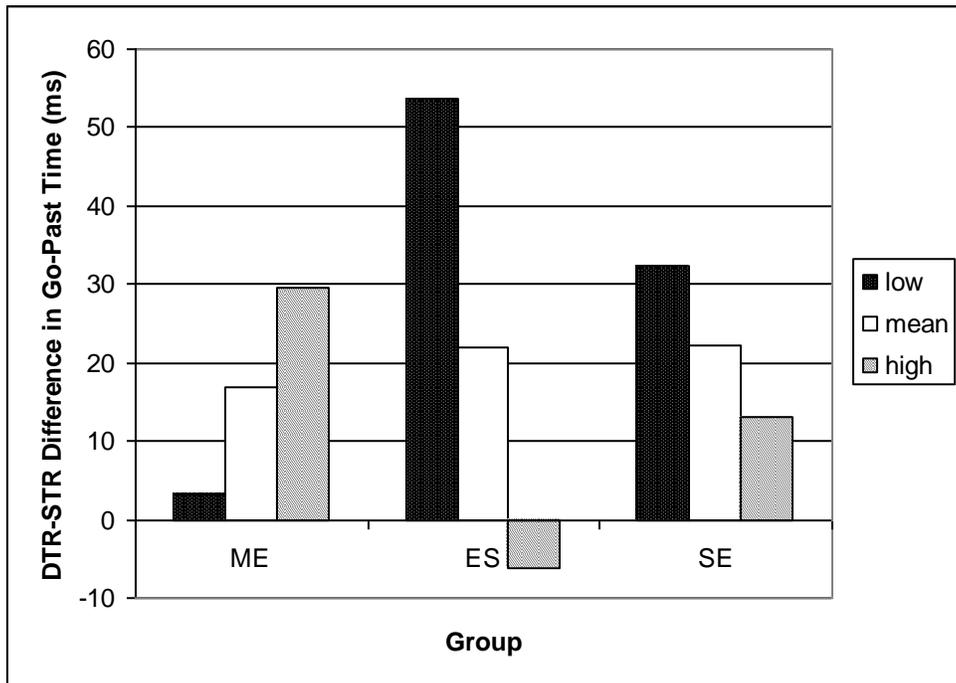
*Gaze Durations:* Gaze durations of the post-target region varied by group, with shorter GD for ES bilinguals compared to ME speakers.

*Go-Past Time:* GPT on the post-target region varied by group, with both ES and SE bilinguals having shorter GPT on this region compared to ME speakers. Critically, the effect for ES bilinguals marginally interacted in a two-way interaction with semantic relatedness, and in a three-way interaction with relatedness and condition (see Figure 7). Whereas ME speakers had longer GPT on the post-target region of DTR sentences than STR sentences when semantic relatedness was high (and less so when it was low), ES bilinguals had longer GPT on DTR sentences when relatedness was low, but exhibited longer GPT on STR sentences when

relatedness was high. As becomes clear in Figure 8, SE bilinguals patterned similarly to ES bilinguals, but the difference between them and the ME group again did not reach significance.

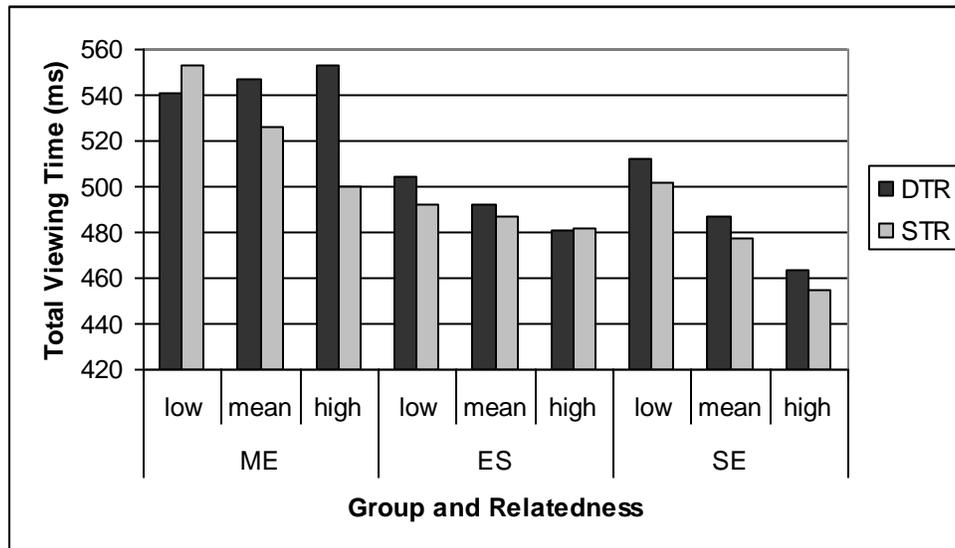


**Figure 7.** Go-past time in the post-target region as a function of group, condition, and semantic relatedness.

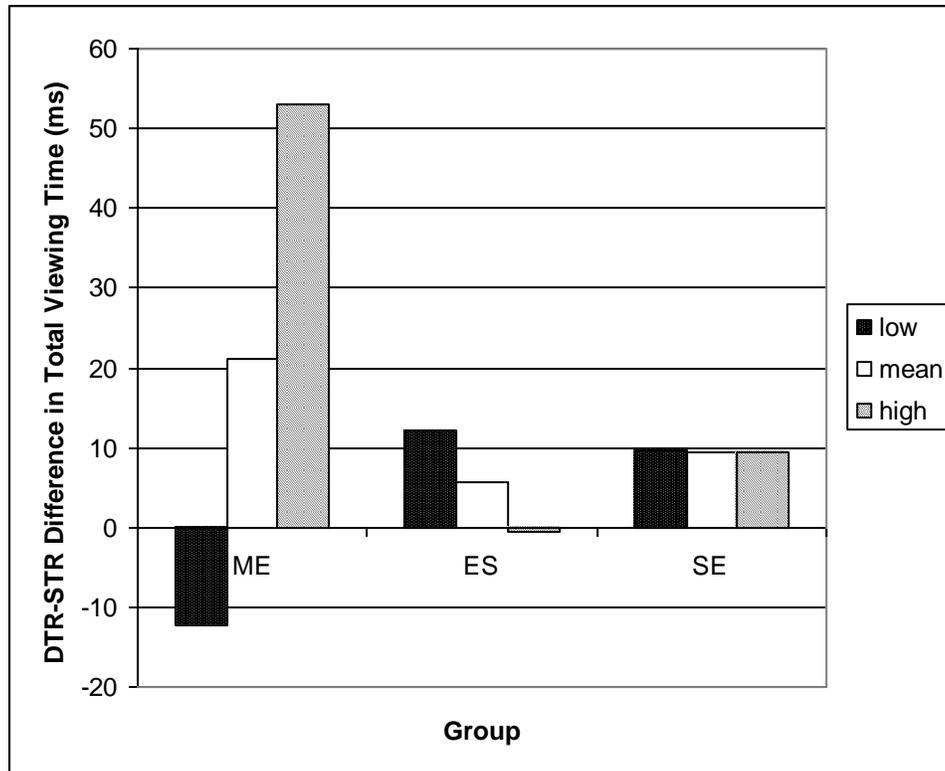


**Figure 8.** Differences in go-past time in the post-target region between DTR and STR sentences as a function of group and semantic relatedness.

*Total Time:* Total viewing time in the post-target region varied by group, such that both bilingual groups spent less time on this region compared to the ME speakers. Importantly, significant two-way interactions between condition and relatedness, and between relatedness and group, were qualified by significant three-way interactions. As can be seen in Figure 9, when semantic relatedness was low, ME speakers showed reduced TT for DTR sentences in comparison to STR sentences, but ES bilinguals showed a reversed pattern with reduced TT on STR sentences compared to DTR sentences. With increased relatedness, this pattern flipped, such that only ME speakers showed reduced TT on STR sentences whereas ES bilinguals showed equivalent durations across the two conditions. The SE bilinguals marginally differed from the ME speakers by showing reduced TT on STR sentences compared to DTR sentences across all levels of relatedness (see Figure 10).

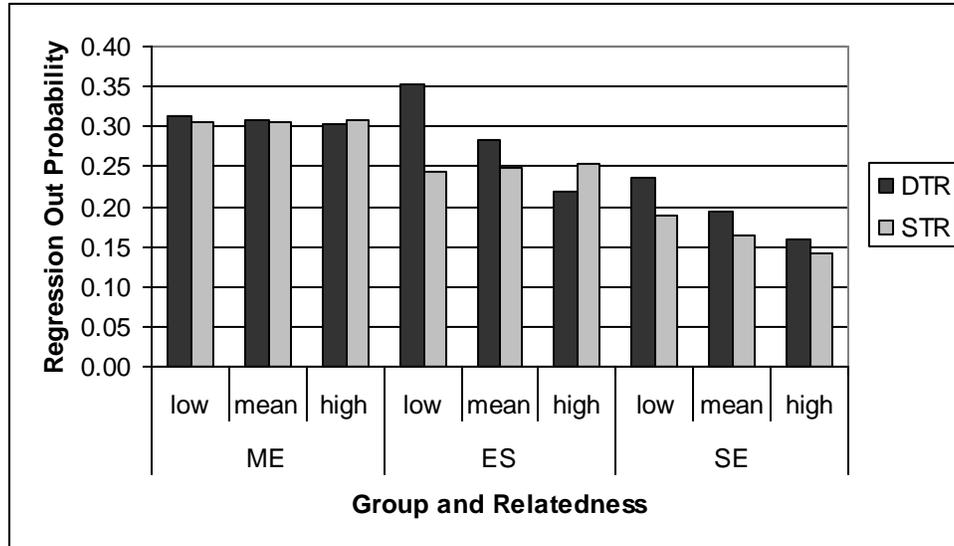


**Figure 9.** Total viewing time in the post-target region as a function of group, condition, and semantic relatedness.

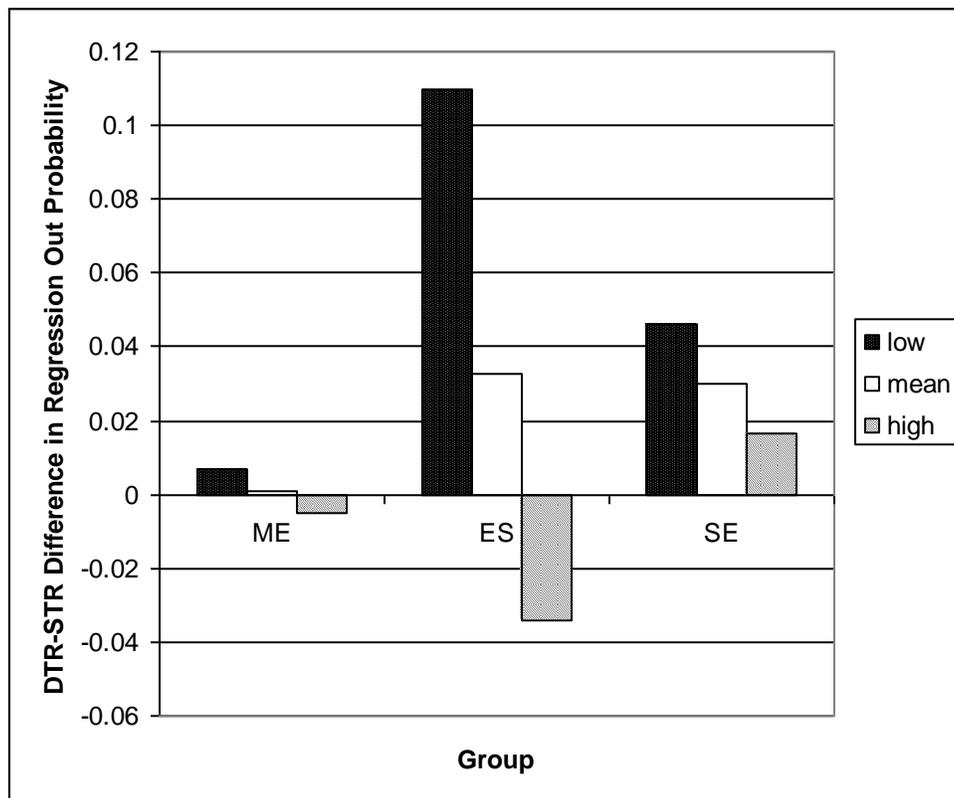


**Figure 10.** Differences in total viewing time in the post-target region between DTR and STR sentences as a function of group and semantic relatedness.

*Regressions Out:* SE bilinguals regressed out of the post-target region less often than ME speakers. Further, the effect of relatedness differed for ES bilinguals compared to ME speakers, but this effect also marginally interacted with condition. As can be seen in Figure 11, ES bilinguals differed dramatically from ME speakers especially when semantic relatedness was low, in that they regressed out of the post-target region in STR sentences less often than in DTR sentences. As semantic relatedness increased, the condition difference decreased, such that when semantic relatedness was high, ES bilinguals regressed out of the post-target region in STR sentences more often. The pattern of means (as seen in Figure 12) suggests that SE bilinguals tended to regress out of the post-target region in STR sentences less often than in DTR sentences, but this effect did not reach significance.



**Figure 11.** Regression-out probability from the post-target region as a function of group, condition, and semantic relatedness.



**Figure 12.** Differences in regression-out probability from the post-target region between DTR and STR sentences as a function of group and semantic relatedness.

*Summary of results in the post-target region:* In this region, four measures indicate that bilinguals differ from monolingual speakers in their processing of STR and DTR sentences. In particular, skipping patterns of the post-target region mirror those of the target region, in that in comparison to ME speakers, ES bilinguals skipped the post-target region more often in DTR sentences when relatedness was low, but more often in STR sentences when relatedness was high. This finding is again consistent with inhibition between unrelated words that share a translation (Elston-Güttler et al, 2005b) as well as increased relatedness between related words that share a translation (possibly due to co-activation; Degani et al., 2011). In later measures on this region, ES bilinguals exhibit a different pattern. Fixation durations (GPT and TT) and regressions out are reduced on unrelated STR sentences compared to DTR sentences, indicating that ES bilinguals find these STR completions easier to process than their controls. This finding is consistent with increased relatedness due to co-activation between (relatively) unrelated words that share a translation (Degani et al., 2011). In contrast, the same ES bilinguals show more RO and longer fixation durations (GPT, and to a lesser extent TT) on STR sentences than DTR sentences when relatedness is high, consistent with interference between highly-related words that share a translation. SE bilinguals differ from ME speakers in TT on the post-target region, and show little difference between STR and DTR sentences as a function of semantic relatedness.

### **2.3.1.6 Summary of main results across regions**

The results reveal several consistent findings across regions. First, the relatedness of the target word to the intended word of the sentence influenced reading behavior in all regions. In the target region, increased relatedness was associated with shorter GD, GPT, and TT, as well as more skips. These effects persisted after sentence naturalness and predictability were partialled

out, suggesting that semantic relatedness is not fully captured by these other covariates. The effect of relatedness further interacted with condition and group, as described below.

Second, reading behavior was partially guided by participants' language background, and these effects were mostly reliable in the post-target region. In particular, in comparison to ME speakers, SE bilinguals regressed out less often from the target and post-target regions, and had shorter GPT and TT on the post-target region. ES bilinguals exhibited shorter GD, GPT, and TT on the post-target region compared to monolinguals. Note that in interpreting these group differences it is especially critical to keep in mind that because behavior on IST sentences was partialled out, these patterns reflect the way in which these participant groups differed on anomalous sentences, in comparison to their baseline reading behavior on non-anomalous sentences. Therefore, it is not the case that SE bilinguals spent less time on the post-target region, but rather that semantic anomaly led to less of an increase in post-target region reading times in this group. As is described below, reading behavior on natural sentences (i.e., IST completions) differed significantly as a function of group.

Third, sentences with DTR completions were associated with more regressions out, but this effect differed for SE bilinguals who showed a reversed pattern. In fact, these overall condition effects are of interest here only to the extent that they are qualified by an interaction with group.

Specifically, of most relevance in the current experiment is the difference between STR and DTR sentences as a function of language background. In the target region, SE bilinguals showed more regressions out on STR sentences compared to DTR sentences. This suggests that SE bilinguals considered STR replacements to be less interchangeable with the IST, and, by extension, less related in meaning to IST compared to DTR sentences. The skipping data in the

same region suggest that although SE bilinguals skipped the target region more often in STR sentences than in DTR sentences when relatedness was low, they did so less often than ME speakers. Similarly, although they skipped the target region more often in DTR than in STR sentences when relatedness was high, they did so to a lesser extent than ME speakers (see Figure 2 and Figure 3). We return to this issue in the Discussion section.

The skipping data further suggest that the condition effect differed for ES bilinguals compared to ME controls, such that when relatedness was low, ES bilinguals skipped the target region in STR sentences less often than in DTR sentences. This finding is consistent with STR sentences creating greater processing difficulty, suggesting that ES bilinguals found low-relatedness STR completions less interchangeable with the target words than DTR completions, and, by extension, less related in meaning to the target words than DTR completions. When relatedness was high, ES bilinguals skipped the target region in STR sentences more often, suggesting that STR completions of high relatedness were easier to process than comparable DTR completions.

In the post-target region, the effects of condition and group consistently varied by relatedness (see Figure 9 above). Focusing on SE bilinguals first, a marginal three-way interaction among condition, relatedness, and group emerged in TT, suggesting that in comparison to ME speakers, SE bilinguals spent less TT on STR completions than DTR completions when relatedness was low. When relatedness was high, SE bilinguals still spent less TT on STR completions, but the difference between STR and DTR sentences was reduced in comparison to that of ME speakers. These findings suggest that when relatedness was low, STR completions were more interchangeable than DTR completions for these bilinguals, but when

relatedness was high, STR completions were not more interchangeable for SE bilinguals than they were for ME speakers.

The ES bilinguals differed from ME speakers in SK, GPT, TT, and RO on the post-target region in the interaction among condition, group, and relatedness. In particular, the probability of skips of the post-target region reveals that ES bilinguals tended to skip the post-target region in DTR sentences more often when relatedness was low, but in STR sentences when relatedness was high. This suggests that by the time the decision to skip the post-target region was made, ES bilinguals considered low-relatedness STR completions to be more difficult than DTR completions, but considered high-relatedness STR completions more interchangeable than DTR completions.

Next, the three-way interactions among condition, group, and relatedness in GPT and TT reveal that ES bilinguals spent more time on the post-target region of DTR sentences compared to STR sentences when relatedness was low. When relatedness was high, ES bilinguals spent slightly more time on STR than DTR completions. Thus, in contrast to the finding observed in the probability of skipping the target and post-target regions, when relatedness was low, STR completions were easier to process than DTR completions. This suggests that by this time point, ES bilinguals found low-relatedness STR completions more interchangeable than DTR completions. When relatedness was high, the difference between DTR and STR completions disappeared for ES bilinguals, but in comparison to ME speakers they spent more time on STR sentences than DTR sentences. Again, in contrast to the skipping data, highly related STR completions were potentially more confusing at the post-target region. Participants' explicit knowledge of the confusability of these items may have started to play a role.

In addition, the RO pattern suggests that ES bilinguals regressed out from the post-target region of DTR sentences more than STR sentences when relatedness was low, but more from STR sentences when relatedness was high. This pattern converges with the duration outcomes on this region, and suggests less difficulty with STR sentences when relatedness is low, and more difficulty with them when relatedness is high. In the Discussion section we return to these findings, and offer an integrative account for the complex interactions of condition, relatedness, and group over time.

Interestingly, the difference between monolingual and bilingual speakers in processing STR and DTR sentences was salient in participant's skipping behavior of the target and post-target regions. The emergence of the effects in skipping probability, as well as the generally high skipping probability across all participant groups, are likely due to the fact that the intended targets in the current experiment were highly predictable (M=50.1%; with 30% of the targets being 90-100% predictable; for similarly high skipping probabilities in sentences with high predictability, see e.g., van Assche et al., 2011). Moreover, many of the target words in the STR and DTR conditions were relatively short (i.e., 3 or 4 letter words; 42%). It is therefore not surprising that participants tended to skip so often.

Skipping probability is typically influenced by form characteristics of the word such as the words' length in number of letters (Brysbaert, Drieghe, & Vitu, 2005; Rayner & McConkie, 1976), but other factors such as the words' predictability have also been shown to independently contribute to skipping behavior (e.g., Rayner et al., 2011). In the EZ reader model of eye movement control during reading (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003) skipping is thought to be initiated after the word is identified and only initially processed, but before semantic processing of the word is completed. That the effects in

the current experiment emerged in the probability of skipping may therefore indicate that lexical (form)-level connections between words that share a translation drive these early translation status effects. As reviewed in the introduction, Elston-Güttler et al. (2005) similarly observed translation status effects in the N200 ERP component thought to reflect lexical (form) level processing (e.g., Bentin, Mouchetant-Rostaing, Girad, Echallier, & Pernier, 1999). They therefore suggested that inhibition at the lexical level (rather than the semantic level) is at the heart of these effects.

### **2.3.1.7 Summary of covariates' effects across regions**

Although the effects of group, condition, and relatedness were of primary importance for evaluating the theoretical frameworks regarding the shared-translation effect, we can examine the effects of stimulus and participant characteristics on the data. First, we observed that performance on IST sentences accounted for significant variability in participants' performance on DTR/STR sentences. Mean IST emerged as a significant predictor across all measures and all regions, suggesting that its inclusion in the model was indeed warranted. We further observed that participants' age accounted for unique variability in the models predicting RO of the target region, and GPT in the post-target region, such that probability of RO and GPT increased with age, consistent with previous findings (e.g., Kliegl, Grabner, Rolfs, & Engbert, 2004).

Second, with respect to item characteristics, our results generally resonate with previous findings, in that higher frequency targets were associated with shorter GD, GPT, and TT on the target region, and increased length was associated with significantly longer GD, GPT, and TT, as well as reduced probability of skipping the target region. Target's part-of-speech had relatively limited influence on reading behavior, with more RO of the target region when the target was a noun compared to when it was a verb, and with longer TT in the post-target region for sentences

with verb targets. The form-similarity of the target word to the intended target word had very limited influence on the eye tracking record, with marginally more RO of the post-target region, and no effects on the target region itself.

Increased target predictability had no influence on the target region, but led to longer FFD and GD on the post-target region. These unexpected findings, especially the absence of an influence for this factor on target reading, are likely due to the reduced range of predictability for the analyzed sentences. Specifically, because sentences with high target-predictability (i.e., IST) were not included, and because item selection intentionally minimized predictability of STR and DTR completions, the range of predictability for STR sentences ranged from 0 to 20%, and was kept at 0 for DTR completions. With other semantic predictors in the model, such as sentence naturalness and the semantic relatedness of the target to the intended target, it is not surprising that predictability exerted relatively little influence.

Sentence naturalness had a stronger effect, such that in the target region reduced naturalness led to marginally longer GD, significantly longer GPT, TT, and more RO. In the post-target region, reduced naturalness led to longer GPT and TT, more RO, but surprisingly also marginally more skips. This latter finding could indicate that on sentences in which naturalness was very low, readers deemed the sentence as not making sense, and skipped the end of the sentence. The fact that comprehension questions followed only natural sentences, as is typically done (e.g., Rayner et al. 2004; Warren & McConnell, 2007; see Warren, 2011, for review), may have contributed further to this reading behavior.

### **2.3.1.8 Group differences in reading natural (IST) sentences**

To examine baseline differences between the groups in reading behavior, we examined the effect of group on mean performance on natural (IST) sentences in each measure and each region.

Table 8 presents means and standard deviations on these measures separated by region. Significant overall effects were probed with t-tests with Bonferroni corrections for multiple comparisons. As can be seen in Table 8, SE bilinguals exhibited longer fixation durations in most regions and measures, and further showed less skipping in the target region in comparison to ME speakers. ES bilinguals exhibited more skipping in all regions, and more regressions out of the post-target region.

**Table 8.** Mean (and SD) performance on IST sentences in each sentence region and measure as a function of group.

Sentence Region	Measure	Group		
		ME	ES	SE
Pre-target	SK	0.08 (0.06)a*	0.17 (0.12)b	0.06 (0.07)a*
	FFD	213.04 (23.47)ab	209.76 (24.22)a	225.32 (33.97)b
	GD	290.72 (44.45)a	268.05 (52.98)a*	327.95 (78.74)b
	GPT	339.48 (74.62)a*	318.99 (70.13)a*	409.62 (134.95)b
	TT	377.56 (83.85)a*	390.91 (98.18)a*	488.70 (156.77)b
	RO	0.11 (0.08)a	0.13 (0.10)a	0.13 (0.09)a
Target	SK	0.16 (0.10)a	0.23 (0.10)b*	0.11 (0.08)c
	FFD	216.18 (23.53)a	212.40 (24.16)a*	235.00 (45.01)b
	GD	246.56 (29.98)a*	236.18 (34.47)a*	291.37 (80.50)b
	GPT	284.79 (46.08)a*	290.11 (59.20)a*	368.53 (176.60)b
	TT	318.71 (46.28)a*	322.30 (68.12)a*	398.05 (126.13)b
	RO	0.09 (0.07)a	0.14 (0.08)a	0.13 (0.11)a

Post-target	SK	0.12 (0.09)ab	0.17 (0.16)a*	0.08 (0.11)b
	FFD	234.91 (33.83)a	234.68 (47.05)a	250.15 (42.37)a
	GD	377.40 (58.23)ab	349.46 (93.92)a*	402.98 (95.97)b
	GPT	486.93 (138.61)a	581.92 (293.44)a	536.13 (175.57)a
	TT	505.46 (97.01)a	494.79 (138.04)a	593.73 (162.62)b*
	RO	0.16 (0.11)a	0.26 (0.21)b*	0.16 (0.13)a

Note: Means in the same row that do not share subscript differ at least at the  $p < .1$  level.

Differences at the  $p < .05$  are marked with a \*.

### 2.3.2 Sentence naturalness ratings

To review, following the eye-tracking session, participants completed a secondary task on the critical sentences. In this off-line task, they rated the naturalness of sentences they had previously seen in the eye-tracking session. Data for one ME participant were unavailable for this task. Analyses were therefore performed on data from 89 participants. Approximately 9.3% of the data (13.9% of the bilingual data) were removed prior to analyses due to unfamiliarity of the shared-translation pair (i.e., the IST, STR, or both). This was done for bilingual participants only, on an item-by-item basis, based on each participant's vocabulary post-test (see Appendix J). Analyses approach and model building were similar to those performed on the eye-tracking measures. For ease of interpretation, the naturalness rating scale was reversed prior to analyses, such that lower ratings indicate that the sentence is perceived as less natural and higher ratings indicate the sentence is perceived as more natural. Table 9 presents the coefficient estimates from these models. In what follows, only significant or marginal fixed effects for the theoretically-relevant predictors of condition, group, and relatedness are reported.

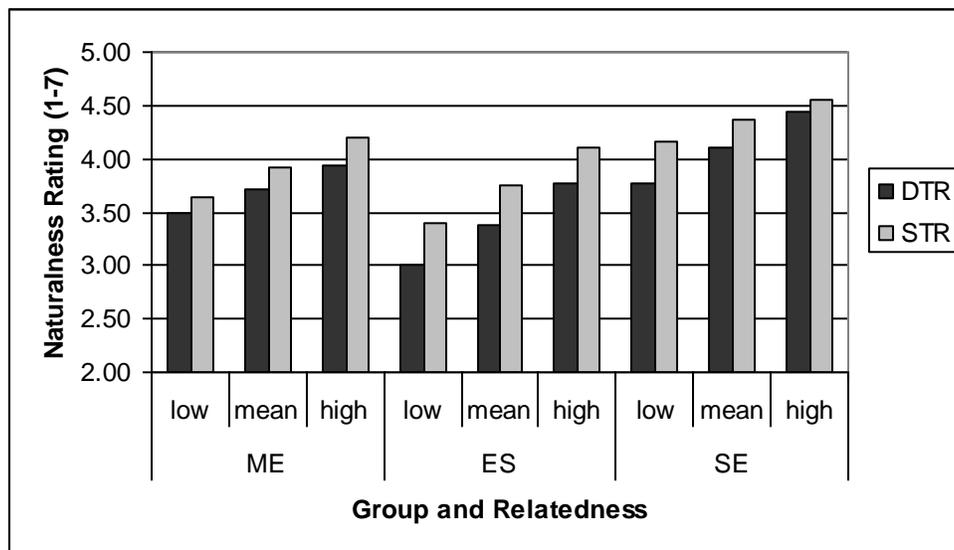
**Table 9.** Coefficient estimates for the sentence naturalness rating task.

	Coefficient
Intercept	3.710**
Participant Age	-0.004
Mean Rating on IST sentences	-0.051
Target Log KF frequency	0.046*
Target Length (in letters)	0.041±
Target Part of Speech [verb]	-0.003
Sentence Predictability Rating	-0.080
Sentence Un-naturalness Rating	-0.694**
Log Form-Similarity to the IST	-0.013
Semantic Relatedness (to the IST)	0.145*
Group [ES]	-0.326
Group [SE]	0.391
Condition [STR]	0.209*
Condition [STR]:Group [ES]	0.155
Condition [STR]:Group [SE]	0.047
Condition [STR]:Semantic Relatedness	0.037
Group [ES]: Semantic Relatedness	0.102±
Group [SE]: Semantic Relatedness	0.075
Condition [STR]: Group [ES]: Semantic Relatedness	-0.051
Condition [STR]: Group [SE]: Semantic Relatedness	-0.124±

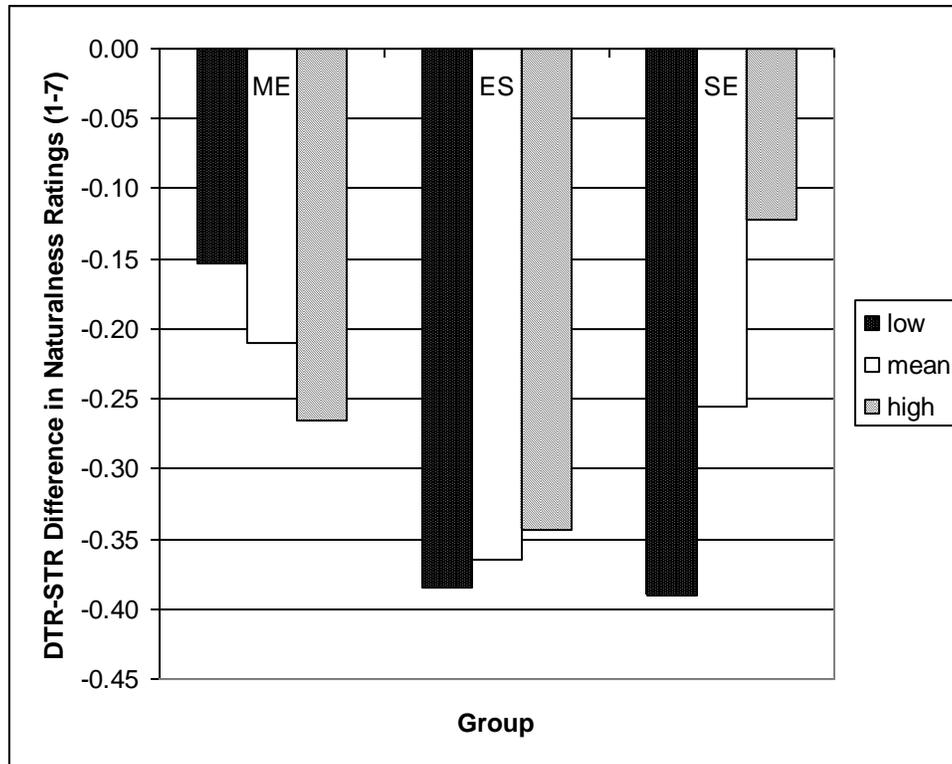
*Note:* ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

As can be seen in Table 9, as the semantic relatedness of the target to the IST increased, sentences were rated as more natural. This effect marginally interacted with group, such that the effect of relatedness was somewhat stronger for ES bilinguals compared to ME speakers. The effect of relatedness further marginally interacted with group and condition as described below.

Sentences with STR completions were rated as more natural than sentences with DTR completions, but this effect marginally interacted with group and relatedness. As can be seen in Figure 13 and Figure 14, all three groups rated STR completions as more natural than DTR completions, but this difference varied by group and relatedness. In comparison to ME speakers, SE bilinguals had a much bigger difference between STR and DTR when relatedness was low, but a smaller difference when relatedness was high. Thus, SE bilinguals found low-relatedness STR completions *more* natural than DTR completions, but in comparison to ME speakers, found high-relatedness STR completions *less* natural than DTR completions.



**Figure 13.** Mean sentence naturalness ratings as a function of group, condition, and semantic relatedness.



**Figure 14.** Mean difference in sentence naturalness ratings between DTR and STR sentences as a function of group and semantic relatedness.

### 2.3.2.1 Group differences in ratings of natural (IST) sentences

The model reported above included a predictor estimating individual's mean rating on natural (IST) sentences. In contrast to the eye tracking data, in which mean performance on IST sentences emerged as a significant predictor in all measures and regions, in the analyses of sentence naturalness it was not significant. Further, the pattern of results remained the same when it was removed from the model.

Nonetheless, mean naturalness ratings on IST sentences significantly varied by group,  $F(2,88)=3.911$ ,  $MSE=.642$ ,  $p=.024$ . T-tests with the Bonferroni corrections reveal that ES bilinguals rated IST sentences as significantly more natural ( $M=6.41$ ) than ME speakers ( $M=5.87$ ), and marginally more natural than SE bilinguals ( $M=5.95$ ).

## 2.4 DISCUSSION

The results of the current experiment clearly demonstrate bidirectional influences in bilingual word meaning. Both SE and ES bilinguals differed from ME speakers in how they processed STRs compared to DTRs when these were embedded in sentence context. Strikingly, the shared-translation effect was more pronounced and reliable in the eye-movement records of ES bilinguals, who exhibited influences of a later-learned L2 on processing L1 sentences. It nonetheless was also present in the performance of SE bilinguals, as indexed by their eye movements and sentence naturalness ratings. The presence of the effects in both groups of bilinguals, in sentence context, extends previous studies that demonstrated bidirectional shared-translation effects for words in isolation (Degani et al. 2011), or that demonstrated the effect in context only in the L1 on L2 direction of influence (e.g., Elston-Güttler et al., 2005b; Elston-Güttler & Williams, 2008). Before discussing the implications of these results, we integrate the findings across sentence regions and dependent measures to provide a detailed account of how the shared-translation effect unfolds over time.

Bilinguals' eye movements and naturalness ratings indicate that replacing a target word in a sentence with its shared-translation counterpart (e.g., replacing 'clock' with 'watch', both sharing the Spanish translation 'reloj') could both ease processing and make it more difficult in comparison to different-translation replacements. The direction of the effect appears to depend on both the baseline relatedness of the words and the time at which processing is sampled. Ease of processing for shared-translation replacements is expected based on previous findings that demonstrated increased semantic relatedness for words that share a translation (Degani et al., 2011; Elston-Güttler & Williams, 2008; Jiang, 2002; 2004). Such increased relatedness has been suggested to stem from strengthened connections due to co-activation in the interconnected

bilingual lexicon (Degani et al., 2001). Increased difficulty in processing shared-translation words has also been documented (Elston-Güttler et al., 2005b), and has been suggested to be the result of lexical inhibitory connections that develop between two (unrelated) words that share a translation, because these serve as mutually-exclusive alternatives in many contexts. The results observed in the current experiment in fact suggest there is merit to both accounts, as described below.

For ES bilinguals, differences emerged across sentence regions and eye-movement measures, and were modulated by the baseline relatedness of the words. To understand how processing may evolve over time, it is useful to examine the reliable effects in order. At the first time point (skips of the target region), ES bilinguals find low relatedness STR completions more difficult (and skip them less often), but high relatedness STR easier (and skip them more often) than DTR completions. This suggests that low-relatedness STR completions are less interchangeable than their controls at this time point, consistent with inhibition between two (unrelated) words that share a translation (Elston-Güttler et al., 2005b). High-relatedness STR completions, in contrast, appear to be more interchangeable than DTR completions, indicating increased relatedness and facilitation between them (consistent with the co-activation account; Degani et al., 2011). At the next time point (skips of the post-target region), ES bilinguals still find low relatedness STR completions more difficult than DTR completions, but high relatedness STR completions easier than DTR completions. Together, these skipping findings are consistent with inhibition (potentially at the lexical level, as suggested by Elston-Güttler et al., 2005b) for unrelated words that share a homonym translation. Related words that share a translation enjoy increased interchangeability for ES bilinguals as compared to ME speakers. At the next time point (GPT and RO of the post-target region) a different pattern emerges. Specifically, low-

relatedness STR completions are easier to process compared to DTR completions, as indexed by both shorter GPT and fewer RO. High-relatedness STR completions, in contrast, are more difficult than DTR completions, as indexed by both longer GPT and more RO. The reversal for highly-related items may be due to the involvement of explicit knowledge, such that because bilinguals may be aware of the potential for confusion on these items they spend slightly more time processing them and verifying their initial interpretation with rereading earlier regions. Finally, TT on the post-target region, which includes second-pass readings of this area, continues to suggest easier processing for low-relatedness STR compared to DTR, and more difficult processing of high-relatedness STR compared to DTR for ES bilinguals.

Thus, the account we put forth posits that the shared-translation effect differs as a function of baseline relatedness, and evolves over time. Unrelated words that share a translation (e.g., ‘doll’ and ‘wrist’ for ‘muñeca’), create increased difficulty earlier in processing, consistent with inhibition at the lexical level (the inhibition account; Elston-Güttler et al., 2005b). This inhibition, however, is soon settled and clears the way for facilitation, reflecting increased semantic relatedness. Related words that share a translation (e.g., ‘watch’ and ‘clock’ for ‘reloj’), exhibit increased relatedness earlier on, consistent with the co-activation account and with previous findings (Degani et al., 2011; Elston-Güttler & Williams, 2008; Jiang 2002; 2004). The advantage for related shared-translation words seem to decay over time, such that explicit knowledge may create interference.

In the eye-movement record, the difference between SE bilinguals and ME speakers in processing STR versus DTR completions was rather weak. It surfaced in the skipping data of the target region, such that SE bilinguals skipped STR more than DTR when relatedness was low (consistent with inhibition for STR sentences), and skipped DTR more than STR sentences when

relatedness was high (consistent with facilitation), but because these effects were dramatically reduced compared to ME speakers, one may claim that they in fact merit the exact opposite interpretation, with facilitation when relatedness is low and inhibition when relatedness is high. These skipping results are therefore difficult to interpret and likely stem from the profound difference in reading style across groups, as will be discussed below. Next, a difference emerged in the regressions out of the target region, with SE bilinguals regressing out from STR sentences more often. This finding seem to suggest that STR completions were less interchangeable with the IST, and by extension less related in meaning than DTR sentences for SE bilinguals. Finally, the marginal effect in total-viewing time of the post-target region suggests that SE bilinguals spent less time on STR than DTR sentences, especially when relatedness was low, in comparison to ME speakers. When relatedness was high they still showed reduced total-time on STR sentences, but the difference was considerably reduced compared to ME speakers (see Figure 10). This suggests that by the post-target region, SE bilinguals, like ES bilinguals, found low-relatedness STR completions more interchangeable than DTR completions, but high-relatedness STR more difficult than their DTR controls.

Interestingly, the results for SE bilinguals from the sentence naturalness task on the same items converge with the end state as reflected in the later eye-movement measures on the post-target region. In particular, in comparison to ME speakers, SE bilinguals rated sentences as more natural when they included a low-relatedness STR completion compared to a low-relatedness DTR completion. In addition, in comparison to ME speakers, they rated high-relatedness STR completions to be less natural than high-relatedness DTR completions. Note that although the difference between ES bilinguals and ME speakers did not reach significance (possibly due to ES bilinguals' higher ratings overall), the pattern of means for ES bilinguals resembled that of SE

bilinguals. The sentence naturalness rating task arguably provides less information on how processing unfolds over time, and reflects a ‘snap shot’ of the end of a process. This end state seems to converge with what we observed for ES (and possibly SE) bilinguals in the later eye-movement measures on the post-target region.

This account highlights two important advantages of the approach taken in the current experiment. First, the current experiment investigated the shared-translation effect as a function of semantic relatedness. In contrast to previous studies that focused either on related words only (e.g., Jiang 2002; 2004) or unrelated words only (e.g., Elston-Güttler et al., 2005b), or that sampled items from the two extremes (Degani et al. 2011), the current experiment included items that varied along the semantic relatedness continuum. Indeed, the effects in question seem to depend on semantic relatedness. The inclusion of items that vary in their baseline relatedness in the same experiment allowed us to propose a comprehensive account of the shared-translation effect, one that does not extrapolate from extreme levels of relatedness.

Second, the eye tracking methodology allowed us to identify how the effect unfolds over time, both with respect to different regions within the sentence, and with different measures within a region. This rich record provides information that is not detectable when sampling participants’ overt responses after completing reading the sentence (e.g., as with the naturalness rating task in the current experiment, or the semantic anomaly paradigm of Elston-Güttler & Williams, 2008).

The findings from the eye-tracking record of SE bilinguals seem to be consistent with the account we proposed above for low-relatedness shared-translation words, namely lexical inhibition that gives way for increased semantic relatedness and facilitation. Note, however, that these results are rather weak and difficult to interpret. In particular, the analyses examining

reading performance on normal (IST) sentences suggest profound differences in reading behavior among the groups. Specifically, SE bilinguals exhibited longer fixation durations across most measures. They further tended to skip the target-region less often than the other groups. These group differences may have obscured the effects of interest, and hinder our ability to make a direct comparison between reading measures of ME speakers and SE bilinguals. Indeed, more systematic investigations of how reading differs in L1 and L2 are needed (see e.g., Frenck-Mestre & Pynte, 1997, for group differences in re-reading measures and potentially in parsing strategies).

Alternatively, reduced cross-language effects for SE bilinguals in the reading task could potentially be explained by proficiency differences. In particular, Elston-Güttler et al., (2005b) observed that low-proficiency German-English bilinguals showed reversed priming in RT and N200 modulation for words that share a German homonym translation. High-proficiency bilinguals did not show any ERP modulation, and further showed no behavioral effects when the words were embedded in sentence context. The authors suggested that increased control allowed these bilinguals to better modulate the activation of the non-target language, such that no influence of the non-target language was detectable. Using the same logic, it may be the case that the SE bilinguals in the current experiment were able to exert a great deal of control over their two languages, such that influences of their L1 were not detectable when they processed their L2. Indeed, in comparison to the ES bilinguals, SE bilinguals had a smaller difference between their L1 and L2 proficiency and use.

The sentence reading and naturalness rating tasks converge in providing clear evidence for cross-language influences across the two languages of bilingual speakers, and suggest that its presence does not depend on the particular pair of languages in question. Rather, the shared-

translation effect is demonstrated here in English/Spanish, a pair of languages that has not been examined before. Moreover, the current experiment demonstrates that the presence of a sentence context does not eliminate cross-language influences (see also Elston-Güttler et al., 2005b; Elston-Güttler & Williams, 2008). Furthermore, to our knowledge, this is the first experiment to demonstrate the manifestation of a bidirectional shared-translation effect in a sentence context. Previous work has demonstrated the influence of an L2 shared translation on L1 words presented in isolation (Degani et al., 2011), but the current work shows that these influences can be detected within an L1 sentence frame. Further, results from Degani et al. (2011) exemplify an L2 influence on L1 for bilinguals who have been immersed in their L2 for 20 years on average. Here, we show that even bilinguals who are tested in an L1 environment exhibit cross-language transfer from L2 to L1 in the form of the shared-translation effect.

The effects observed in the current experiment take the form of a divergence of bilingual speakers from monolingual controls. The difference between STR and DTR completions is not observed only for bilingual speakers, but rather takes a different form in the performance of bilingual versus monolingual speakers. Several reasons can explain why monolinguals do not show equivalent performance on STR and DTR completions. First, different words are used as DTR and STR completions. Although we made every effort to equate these words on important lexical characteristics (see Appendix A), residual difference may still be present. Furthermore, although we controlled for relatedness differences between STR and DTR completions to the IST by attempting to match these words in isolation and by co-varying this factor in the model, we did not control for differences in relatedness when the words are embedded in context. Thus, a particular completion may instantiate a more similar concept to the original sentence than another, despite comparable relatedness in isolation. For instance, the shared-translation pair

‘dress-wear’ and its different-translation control ‘dress-hang’ received comparable relatedness ratings in isolation (3.05 and 3.06, respectively). However, in a sentence context ‘For Halloween, he was planning to dress as Count Dracula’, the STR ‘wear’ perhaps instantiates a more similar meaning than the DTR ‘hang’. Note that across items some of these differences are likely captured by the sentence naturalness ratings, but some systematic differences may nonetheless remain. These uncontrolled differences may be the reason monolingual English speakers show differences in their reading behavior in STR and DTR sentences. Critically, bilinguals of English and Spanish diverge from this behavior and show patterns that are in line with the translation status of the word. Future studies could nonetheless attempt to better control differences in semantic relatedness across condition completions when the words are embedded in sentences.

A potentially important factor that was not examined in the current experiment is translation probability, or dominance. In particular, when two English translations are available for a given shared translation in Spanish, one of them may be more frequently given as a translation. Prior et al. (2007) examined such translation probability of multiple translation words, and found that when translating from L1 to L2, less-proficient bilinguals tended to produce the less probable translation in comparison to more proficient bilinguals. Laxén and Lavuar (2010) further showed that translation recognition of translation-ambiguous words differed as a function of translation probability. The more dominant translation was recognized more quickly and accurately than the non-dominant translation (see also Eddington & Tokowicz, 2011). These findings suggest that one of the translations of a shared-translation word (e.g., ‘watch’ or ‘clock’) may be linked more strongly to the shared Spanish word (e.g., ‘reloj’). Elston-Güttler et al. (2005b) examined if the shared-translation effect is modulated by dominance, and observed somewhat stronger effects when the dominant meanings served as the

targets. Note, however, that dominance in that study was determined based on meaning dominance of the German homonyms, rather than translation dominance per se.

This brings us to an important issue in the interpretation of the shared-translation effect. We have proposed that the shared-translation effect stems from changes that occur in the semantic relatedness of words (see also Degani et al., 2011). The meanings associated with two words that share a translation become more interconnected than they would have had they not shared a label in the bilingual's other language, or for monolingual speakers. Others have interpreted similar effects under a processing framework, and suggested that such effects reflect online activation of the non-target translation (e.g., Elston-Güttler et al., 2005b; see also Thierry & Wu, 2007). The results of the current experiment do not allow differentiation of these two accounts because the findings may reflect a stable change in semantic representations or activation of the non-target (Spanish) language while bilinguals perform the task. Nonetheless, Marian and Kaushanskaya (2007) suggested that instances of semantic transfer in production of autobiographical memories, which are similar to the effects we present here, are more likely due to changes in representations than to activation of the non-target language, because their frequency depended less on the language environment in which the retrieved memories were initially encoded. Borrowing of words from the non-target language, in contrast, seems to be more transient and to depend on language activation.

There are several ways in which future studies may attempt to discern if the shared-translation effects demonstrated here reflect a stable change in semantic representations or non-target language activation. First, translation dominance is expected to exert a strong influence on the manifestation of the effect under the activation account. Specifically, the ease with which a particular translation activates the shared-translation word should vary with the translation's

probability. If, however, the shared-translation effect reflects a stable change in the semantic representation of the two translations, the online ease of activation of the shared translation is less relevant. A second approach to differentiate a representation from a processing account is to manipulate the time constraints or SOA of a given task. Such a manipulation should exert an influence on non-target language activation but should not affect stable representational differences. If the shared-translation effect is due to online activation of the shared-translation word, then reducing the time available for participants to perform the task may make such activation less likely (for a similar discussion, see Morford et al., 2011). Finally, one can modulate the language context in which the task is performed to differentiate representation- from activation-type accounts. Although there is clear evidence for non-target activation even in a single language context (e.g., van Hell & Dijkstra, 2002), language context may nonetheless modulate the strength of non-target activation (see Elston-Güttler, Gunter, & Kotz, 2005a; for discussion, see Degani & Tokowicz, 2010b). Ongoing research takes advantage of this approach by testing bilinguals in both languages, and manipulating language order. Stronger cross-language influences are expected in the second block if the shared-translation effect reflects activation of the non-target language rather than a (stable) change in semantic representation.

The presence of cross-language transfer in both languages serves to support the interconnected nature of the bilingual lexicon. In particular, one language may exert an influence on processing of the other only if these languages are highly interconnected in the bilingual lexicon. Moreover, as discussed in Degani et al. (2011), the influence of a shared L2 translation suggests that semantic representations are subject to influences from both L1 and L2 words. If, as proposed by Jiang (2000; 2004), L2 words access a copy of the L1 meaning, then the meanings of words in L2 have no means by which to influence semantic processing of L1 words.

If in contrast, both languages are linked to the semantic representations, then bidirectional patterns of cross-language semantic influences are to be expected. The results of the current experiment clearly support this latter architecture, one in which words from both languages are interconnected, and are both linked to meaning representations.

To summarize, the present experiment demonstrated bidirectional cross-language influences for bilingual speakers when they process the meaning of words embedded in context. Further, the results suggest that the processing of words that are linked via a shared translation in another language known to a bilingual may be subject to both inhibition and facilitation. We propose that unrelated words that share a translation may be inhibitorily connected at the lexical level, leading to initial inhibition of unrelated shared-translation words, but this lexical inhibition is complemented with increased relatedness at the semantic level, leading to facilitation at later stages of processing. Related words that share a translation are initially facilitated, but explicit knowledge may nonetheless increase interference in processing these words under some task conditions. Indeed, the current experiment intentionally created contexts in which even the highly related shared translation counterpart is inappropriate and creates an anomaly.

The present experiment demonstrated that language differences in the mapping of words to meanings carry consequences for the semantic interpretation of words. Further, these consequences of translation ambiguity are bidirectional in nature, such that learning an L2 influences the way bilinguals read sentences in their L1. As such, the results of the current experiment contribute to the growing body of research demonstrating the dynamic nature of the lexicon, and the ability of language experience to continue to guide language processing even in adulthood.

### 3.0 EXPERIMENT 2: THE SPLIT-TRANSLATION EFFECT IN JUDGMENTS OF SEMANTIC RELATEDNESS

#### 3.1 INTRODUCTION

Words are notoriously ambiguous in meaning. A single word can refer to multiple slightly different referents, or even to completely different referents, in different contexts. The word ‘beam’ for instance, can refer to a wooden beam in the context of gymnastics, but to a laser beam in the context of physics. The different senses of words may be more or less related in meaning, and may share some semantic features. For instance, both senses of the polysemous word ‘beam’ encompass a referent with a straight line. In other cases, a word can encompass two unrelated meanings, for which it is more difficult to identify a shared set of semantic features (e.g., the homonym ‘bark’, referring to the sound a dog makes or to the outer layer of a tree). Such homonyms are typically thought to have been accidentally created in the language, such that two separate lexical entries happen to share form (e.g., Klein & Murphy, 2001, 2002)

Words can therefore vary in the *degree of relatedness* of their different nuances of meaning, which we will refer to as intra-word sense relatedness. The present experiment examined whether this degree of relatedness would be influenced by the translation status of the ambiguous word in a bilingual’s other language. In particular, we examined whether the two senses of an ambiguous word may be more (or less) related when a single word also captures

these two senses in a bilingual's other language. For example, the word 'operación' in Spanish refers to both the military and the mathematical senses of the English word 'operation'. In contrast, each sense of the word 'ring' is translated into a different word in Spanish; diamond ring corresponds to 'anillo' whereas a loud ring corresponds to 'timbre'. Here, we asked if two senses with a shared translation in Spanish ('joint translation' condition) differ from two senses with independent translations in Spanish ('split translation' condition) for bilinguals of Spanish and English, compared to monolingual English speakers.

An influence of a bilingual's other language on the degree of relatedness of intra-word senses is important to examine for several reasons. First, similar to the shared-translation effect (see Chapter 2; Degani et al., 2011), such cross-linguistic influences, and especially those from L2 on L1, highlight the dynamic nature of the bilingual lexicon and exemplify the interconnectivity between languages of multilingual speakers. Second, as reviewed below, relatedness of meanings of ambiguous words influences how ambiguous words are processed in and out of context. Thus, if knowledge of another language exerts an influence on the degree of relatedness, bilinguals may no longer process ambiguous words in the same way monolingual speakers do. Third, an influence of a later-learned L2 on the degree of relatedness of sense of an L1 word would suggest that learning, in general, can lead to changes in semantic representations. Other learning experiences, such as domain expertise, can be similarly hypothesized to change semantic representations. To illustrate, as one learns distinctions and nuances of a particular domain, two concepts may become more differentiated. Similarly, when a person learns the underlying mechanism of a particular phenomenon, he/she may come to realize the fundamental similarities of two concepts. This could apply to two senses or meanings as well. For instance, the Hebrew word 'kesef' refers to both 'silver' and 'money'. When one learns that silver was

once used for money exchange, the perceived relatedness of these two senses may change. Thus, one can envision a change in the relatedness of two concepts (regardless of whether they map onto one or two lexical forms) as a function of learning.

Two general questions were addressed in the current experiment. The first pertains to the influence of L1 on L2, and concerns the degree to which intra-word sense relatedness is organized or guided by language background. This is a ‘cross-linguistic’ comparison in that we compare processing of English words by native Spanish speakers (Spanish-English bilinguals) and native English speakers (monolingual English speakers). The second question pertains to the dynamic nature of intra-word senses more generally, and focuses on the degree to which intra-word sense relatedness can undergo change. We focus here on the influence of another language on these internal structures, but this can be taken as an example of within-individual changes that result from learning. We examine whether learning Spanish as an L2 can lead to changes in processing of L1 English. Rather than tracking individuals over time, we compare two groups of native English speakers cross-sectionally, who have or have not learned Spanish as an L2 (English-Spanish bilinguals vs. monolingual English speakers). The influence of L2 on L1 is thus relevant in constraining the nature of the bilingual lexicon, and more broadly in demonstrating changes in intra-word sense relatedness as a function of learning.

In the remainder of the introduction we briefly review how semantic ambiguity influences language processing, and focus on the importance of semantic relatedness for processing of ambiguous words. We then consider inhibition and facilitation in the connections among intra-word senses, before outlining the predictions for the current experiment.

### 3.1.1 Semantic relatedness effects in ambiguity processing

Semantic ambiguity has been studied extensively not only because of its prevalence, but also because it allows controlled examination of the role of context in lexical access. If context alone guides lexical access, then embedding ambiguous words in context (e.g., ‘When you need cash you should simply go to the *bank*’) should eliminate activation of the contextually-inappropriate (e.g., river bank) meaning of the word (e.g., Schvaneveldt, Meyer, & Becker, 1976). Alternatively, if lexical access initially occurs irrespective of context, then both meanings of the ambiguous word should receive bottom-up activation, such that the effect of context comes into play only later in processing (e.g., Hogaboam & Perfetti, 1975; Onifer & Swinney, 1981). This theoretical debate has sparked numerous investigations comparing ambiguous and unambiguous words, in and out of context. It is now generally accepted that both context and frequency of meaning (i.e., dominance) of the ambiguous word interact during lexical comprehension (e.g., Duffy, Morris, & Rayner, 1988; see Degani & Tokowicz, 2010b, for the role of these factors in cross-language ambiguity).

One key finding that emerged from this extensive line of research is a difference in processing between ambiguous and unambiguous words. In particular, ambiguous words were initially found to be processed more quickly than unambiguous words in lexical decision tasks (e.g., Kellas, Ferraro, & Simpson, 1988). This so called ‘ambiguity advantage’ informed theoretical and computational models describing the course of lexical comprehension, and led researchers to suggest that activation of multiple entries in the lexicon facilitates processing (for review, see, e.g., Klepousniotou & Baum, 2007). However, more recent studies have highlighted the relevance of the degree of relatedness of the different senses/meanings of ambiguous words, and claimed that this dimension had generally been overlooked by most previous research (e.g.,

Rodd, Gaskell, & Marslen-Wilson, 2002; see also Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007). In particular, Rodd and colleagues showed that lexical decisions to words with multiple related senses are faster relative to words with fewer senses, but that lexical decisions to ones with multiple unrelated meanings are in fact slower (see also Armstrong & Plaut, 2008; Klepousniotou & Baum, 2007; see Berretta et al., 2005, for converging MEG results). They further showed that relatedness of senses accounted for unique variability in the processing of ambiguous words, such that ambiguous words were processed more quickly as the relatedness of their senses increased.

The empirical evidence shows a processing difference between homonyms (with unrelated meanings) and polysemous words (with related senses). This difference was taken to suggest a potential qualitative difference in the representation of these word types; whereas there is general agreement that different meanings of homonyms are represented with separate lexical entries, the representation of senses of polysemous words in single versus multiple entries continues to be debated. To investigate whether ambiguous words are represented with a single or multiple entries, the word is often presented twice and repetition priming is measured. Critically, the word could be repeated instantiating the same or a different sense. Masson and Freedman (1990) showed that same-sense repetition led to stronger priming in lexical decision compared to different-sense repetition, and suggested that different senses are therefore represented in separate lexical entries. Using a similar logic, Klein and Murphy (2001) asked participants to make timed judgments on whether a phrase biasing a particular sense of the ambiguous word (e.g., wrapping paper) made sense (henceforth, this task will be referred to as a sensicality judgment task). Critically, the same word was repeated later in a phrase instantiating either the same (e.g., shredded paper) or different (e.g., daily paper) sense of the ambiguous

word. Same-sense repetition led to more accurate responses than different-sense repetition, and they concluded that the different senses of ambiguous words are functionally distinct. Interestingly, Klein and Murphy examined polysemous words separately from homonyms, but found a similar pattern for both types of words, suggesting that both are represented with multiple entries. In a second study, Klein and Murphy (2002) used a categorization task in which participants were asked to categorize a target phrase (e.g., wrapping paper) with either a polysemous choice (e.g., daily paper), a taxonomic choice (e.g., a smooth cloth), or a thematic choice (e.g., sharp scissors). Same-sense choices were categorized together more often than taxonomic or thematic choices, but different-sense choices were categorized together less often than taxonomic or thematic choices. Interestingly, the similarity of the different-sense phrases correlated with how often they were categorized together, such that the more similar the two senses were, the more likely they were to be categorized together.

The conclusion from these studies (Klein & Murphy, 2001; 2002; Masson & Freedman, 1990) was that different senses of polysemous words are represented separately in the lexicon, and they are thus comparable to homonyms with unrelated meanings. Other studies did not support this conclusion. For example, Klepousniotou, Titone, and Romero (2008) also used the timed sensicality judgment task (Klein & Murphy, 2001) and compared processing of same- or different-sense repetitions for ambiguous words with highly overlapping, moderately overlapping, and low overlapping senses. Consistent with the original Klein and Murphy (2001) study, processing of same-sense repetition was facilitated relative to neutral (/\*\*\*\*/) preceding the ambiguous word) regardless of the degree of sense overlap. Different-sense repetition differed, however, from same-sense repetition for words with low- and moderately-overlapping senses, but not for words with highly-overlapping senses. The authors took their results to

suggest that homonyms, with separate lexical entries, are processed differently than polysemous words with highly overlapping senses, presumably because the latter are represented with a unified lexical entry.

A recent MEG investigation further supports this dissociation. Pylkkänen, Llinás, and Murphy (2006) had participants perform timed sensicality judgments (Klein & Murphy, 2001) to target phrases (e.g., liberal paper) primed with related or unrelated phrases, in one of three conditions: homonym, polysemous, and semantic relations. In the homonymous and polysemous conditions the related prime biased a different meaning/sense of the word (e.g., lined paper), and in the semantic condition the target phrase was replaced with a synonym (e.g., monthly magazine), such that the semantic relatedness of the different senses was maintained. Priming was observed in all three conditions for related versus unrelated primes, but the MEG record elucidated slightly different processing for the three types of words. Focusing on the M350 component, typically thought to reflect initial stages of lexical and morphological access (e.g., Beretta et al., 2005), related (different-sense) primes resulted in delayed *left* M350 latencies for homonyms but not for polysemous words. Polysemous repetition was associated with a less-typical *right* lateralized temporal activity in the 300-400 ms window in the opposite direction of that observed in the semantic condition. These results are relevant in the dissociation they demonstrate between homonyms and polysemous words, suggesting representational differences.

The question of representation of polysemous words is to some extent separate from the question asked in the current investigation, namely whether changes to the representation can occur as a result of knowledge of another language. Even if polysemous words are represented qualitatively differently from homonyms, some gradient of semantic relatedness of polysemous words should be considered, because there are polysemous words with highly overlapping senses

(e.g., oil, canola oil – engine oil) and polysemous words with less related senses (e.g., beam, laser beam – wood beam). A change in intra-word sense relatedness can be conceptualized as occurring when the different senses are represented by separate lexical entries, by assuming a change in connection weight between these entries, or in the distance between them in the semantic space. Conversely, under the single lexical entry hypothesis, a change may entail increased cohesiveness of the representation, having the senses cluster together more. The question we ask here is whether these differences in semantic relatedness are consistent with the mapping of words to meanings in another language.

Moreover, although some researchers posit a qualitative difference in representation between homonyms and polysemous words (but see Klein & Murphy, 2001), the practical distinction between these two types of words is challenging (see e.g., Durkin & Manning, 1989, for discussion). It is arguably more useful to treat intra-word sense relatedness as a continuum. Indeed, as mentioned above, Rodd et al. (2002) obtained relatedness ratings for their stimuli and found that this factor was a positive predictor of performance for ambiguous words. Similarly, Klepousniotou and colleagues (Klepousniotou & Baum, 2007; Klepousniotou et al., 2008) examined a relatedness continuum for intra-word senses of polysemous words. In an auditory lexical decision task (Klepousniotou & Baum, 2007), homonyms with unrelated senses differed from both polysemous words with very highly related senses (metonymy, e.g., ‘rabbit’, referring to the animal and to the meat) and polysemous words with less related senses (metaphor, e.g., ‘lip’, referring to the organ of the body and to the edge of a vessel). In a visual variant of the task the two types of polysemous words (metonymy and metaphor relations) differed from each other. Thus, even if polysemous words are qualitatively different from homonyms, a gradient of intra-

word sense relatedness should be considered, and this gradient could be hypothesized to be partially influenced by a bilinguals' other language.

### **3.1.2 Inhibition and facilitation among intra-word senses**

In the current experiment, we examine if a shared lexical form can lead to changes in intra-word sense relatedness. One can hypothesize that a shared lexical form would cause two meanings to become more related as a function of co-activation (Degani et al., 2011). In particular, as reviewed in Chapter 2, Degani et al. (2011) observed that bilinguals but not monolinguals rated two words that share a translation in a bilingual's other language as more similar in meaning than two words with different translations. To explain this finding they proposed that two meanings/translations that share a label are co-activated whenever that shared label is encountered. Based on Hebbian principles (Hebb, 1949), this co-activation leads to strengthened connections, such that the two meanings (and translations) become more related in meaning than they would have been had they not shared a label. Alternatively, one can postulate that two meanings that share a lexical form inhibit each other because they are mutually exclusive in any given context (e.g., Chwilla & Kolk, 2003; Elston-Güttler et al., 2005b). For example, because only the financial meaning of the word 'bank' is appropriate when a person intends to withdraw cash, and only the riverside meaning is appropriate when one goes fishing, these two meanings become less related due to mutual inhibition.

It is challenging to contrast these two predictions empirically within a language because it is not clear how to establish the correct baseline. How does one estimate the relatedness of the sound a dog makes and the outer layer of a tree, had they not shared the label 'bark'? One approach could be to use synonyms. For instance, one can compare the relatedness of phrases

instantiating two meanings of an ambiguous word (e.g., lined paper – liberal paper) to phrases where one of these was replaced with a synonym (e.g., lined paper – monthly magazine) (e.g., Pylkkänen et al., 2006). One limitation of this approach, however, is that exact synonyms are rather scarce, if any even exist (for discussion see, e.g., Edmonds & Hirst, 2002; Inkpen & Hirst, 2006). Moreover, if one assumes the two meanings undergo change as a result of a shared lexical form, this change may carry over to the relatedness with the synonymous phrase.

An alternative approach, adopted in the current experiment, is to compare two meanings that share a label in one language (English; split condition) to two meanings that share a label in two languages (English and Spanish; joint condition). We can establish matching in baseline relatedness between these pairs of meanings with monolingual speakers. A difference between these conditions could be attributed to the effect of an *additional* shared label in the joint condition for bilingual speakers.

Such an approach can help examine if intra-word senses are more or less related due to their shared label. The extensive literature on ambiguity processing suggests that the two meanings of ambiguous words may become more related because both are initially activated (though this activation is partially a function of the meaning dominance and contextual constraint, e.g., Simpson, 1981). This initial co-activation may lead to increased relatedness of the different meanings of ambiguous words. This is especially likely for polysemous words, for which multiple senses could be relevant in the same context. For example, in the sentence ‘your book is not only badly written, it is too heavy’, the physical sense of the word book and the novel meaning are both relevant (for discussion, see Klein & Murphy, 2001, p. 273).

Following this initial co-activation, the meaning that is appropriate in a given context is selected, but the fate of the inappropriate meaning is in question. Its activation either simply

decays to baseline or it is actively suppressed or inhibited below baseline. If the latter option is true, then one can postulate that inhibitory connections are likely to develop between the two alternative meanings of the ambiguous word. Several studies provide evidence relevant to this issue. For example, Chwilla and Kolk (2003) examined the relationship between intra-word meanings/senses by using a double word priming procedure with ERPs. Specifically, they compared priming for conditions in which a target word was preceded by one related prime and one unrelated prime (e.g., kidney-soda-organ or soda-kidney-organ) to a condition in which the target was preceded by two primes that were related to different meanings of the target and were unrelated to each other (e.g., kidney-piano-organ). Results from a lexical decision task showed additive effects for the two related primes in both reaction time and the N400 mean amplitude, suggesting neither inhibition nor facilitation between the two meanings of the ambiguous word. In a relatedness judgment task of the prime(s) to the target, however, the results showed under-additive priming of the two related primes. This indicates that one or both primes produced less facilitation than when presented alone, suggesting inhibition between the different meanings of the ambiguous words.

It is important to note that this inhibition between multiple meanings of ambiguous words was observed only when the task required meaning selection (i.e., in the relatedness judgment task) but was absent in the lexical decision task (see also Balota & Paul, 1996). It is therefore not clear whether such inhibition leads to stable changes in semantic representation of the type we are investigating here. Furthermore, these studies presumably focused on ambiguous words with unrelated meanings, but the degree of relatedness of the different meanings was not examined closely.

Evidence regarding facilitation and inhibition between different senses of polysemous words comes from one of the experiments (Experiment 5) in the Klein and Murphy (2001) study described earlier. Specifically, they examined if the advantage for same-sense repetitions over different-sense repetitions in their sensicality judgment task was due to facilitation of the same sense, or due to inhibition between different senses of polysemous words, or both. Focusing on polysemous words only, they contrasted same-sense repetitions (daily paper – liberal paper) and different-sense repetitions (wrapping paper – liberal paper) with a neutral condition (\_\_\_\_ paper – liberal paper). Their results provided support for both facilitation of same-sense repetition and inhibition of different-senses, in that the neutral condition elicited slower and less accurate decisions compared to the same-sense repetition, but faster and more accurate decisions relative to different-sense repetition.

Notably, however, Klepousniotou et al. (2008) did not observe inhibition for different-sense repetition using a similar paradigm with a slightly different neutral condition (‘\*\*\*\*\* paper’). The results from the study by Masson and Freedman (1990) also suggest no inhibition. In particular, they examined repetition effects and demonstrated that repeating an ambiguous word in a context instantiating the same sense facilitated lexical decision times compared to a non-repeated condition, but that repeating the word in a context instantiating a different sense did not produce significant priming. In a naming task, repetition in both same-sense and different-sense contexts yielded significant priming relative to no repetition. These results suggest that different senses of ambiguous words do not suppress each other because such suppression would have resulted in negative priming for different-sense repetition.

To summarize, it is generally agreed that both meanings of ambiguous words receive initial activation when encountered, even in context (for a review, see, e.g., Gorfein, 2001).

However, whether this co-activation is followed by inhibition between different senses/meanings of ambiguous words is still an open question.

### **3.1.3 The current experiment**

In the current experiment we examined ambiguous words that vary in the relatedness of their senses, as determined by the ratings of monolingual English speakers (see Appendix F). Thus, rather than dichotomizing homonyms from polysemous words (e.g., by counting the number of entries the word corresponds to in a dictionary) we chose to rely on the psychological relatedness of the different senses/meanings (for a discussion of the consistency of these two measures, see Rodd et al., 2002). Previous researchers have pointed out the difficulty in clearly distinguishing homonyms from polysemous words, and some have similarly adopted a continuous measure to capture intra-word sense relatedness (e.g., Klepousniotou et al., 2008; Rodd et al., 2002).

Each ambiguous word was embedded in two expressions, such that each expression highlighted a different sense of the ambiguous word (e.g., expiration date-dinner date). These pairs of expressions were then presented together, and participants were asked to make a timed relatedness judgment ('yes' or 'no') on each pair. Critically, some pairs of expressions instantiated two senses that are captured by one word in Spanish ('joint translation' condition) and some pairs instantiated senses that correspond to different Spanish translations of the ambiguous word ('split translation' condition). For each language group (i.e., monolingual English, English-Spanish bilinguals, and Spanish-English bilinguals) we compared relatedness judgments to joint and split translation expressions.

For all participant groups, relatedness ratings are predicted to influence online relatedness judgments (e.g., Jiang, 2002). Specifically, pairs of expressions that were rated in a norming task

as more similar in meaning should be more likely to elicit a related ('yes') response in the online judgment task. Further, for trials in which a 'yes' response is made, decisions should be faster for more-related pairs compared to less-related pairs (e.g., Thierry & Wu, 2007). Conversely, for trials in which an unrelated ('no') response is made, decision times should be slower for more-related pairs compared to less-related pairs (e.g., Morford et al., 2011). This pattern would strengthen our confidence that the task indeed reflects participants' semantic relatedness judgments.

Of interest in the current investigation is the difference between joint- and split-translation expressions for bilinguals compared to monolinguals. Two alternative predictions can be contrasted for this effect. Increased relatedness for joint-translation expressions would be expected if one assumes that a shared label leads to increased connectivity based on Hebbian principles (Hebb, 1949). This account has been suggested to explain the shared-translation effect (Degani et al., 2011; see also Chapter 2) in which two words that share a label in a bilinguals' other language were perceived by bilinguals as more similar in meaning than two words with different translations. This account posits that co-activation of the shared translation and its two senses results in stronger semantic links between the two senses. Senses in the split-translation condition (e.g., dinner date – expiration date) are linked to a shared label in English only, whereas senses in the joint-translation condition (e.g., military operation – mathematical operation) are linked to a shared label in Spanish ('operación') as well. Thus, in the 'split translation' condition the two senses will be co-activated only when the English word is encountered, but in the 'joint-translation' condition the two senses will be co-activated whenever either the English word or the Spanish word are encountered. The co-activation account was suggested to apply to shared-translation words regardless of their baseline semantic relatedness.

By extension, two senses will grow more similar in meaning by their shared Spanish translation, irrespective of their initial relatedness. Therefore, according to this co-activation account, joint-translation expressions are predicted to be more similar in meaning than split-translation expressions for homonymous and polysemous words alike.

Alternatively, two concepts or words that share a label may be predicted to be *less* similar in meaning because of inhibition. In particular, because the two meanings of an ambiguous word are mutually exclusive in most contexts (e.g., the river edge meaning of the word ‘bank’ is inappropriate in a financial context), they are hypothesized to inhibit each other, such that the activation of one meaning would reduce activation of the other meaning. This account has been suggested to explain reduced priming for two English words that share a German homonym translation (Elston-Güttler et al., 2005b), although the authors focused on lexical rather than semantic level inhibition. As reviewed above, the empirical evidence regarding this intra-word sense/meaning inhibition is inconsistent. Moreover, inhibition might be hypothesized to play a role for unrelated meanings (of homonyms), but it is less clear why related senses (those of highly overlapping polysemous words) would inhibit each other, given that they sometimes fit the same context (for discussion, see Degani et al, 2011; Klein & Murphy, 2001). Here, the inhibition account predicts that two senses that share a label in English and in Spanish (i.e., joint-translation condition) would be less related than two senses that share a label in English but correspond to two translations in Spanish (i.e., split-translation condition), especially if the two senses are generally unrelated in meaning.

To examine these issues, we compared semantic relatedness judgments of (English-Spanish and Spanish-English) bilinguals and English monolinguals to pairs of expressions containing an ambiguous word (e.g., diamond ring – loud ring) that either share a Spanish

translation (joint translation) or correspond to two different Spanish translations (split translation). Because semantic relatedness judgments are subjective by nature, individuals may differ in the criterion they impose for relatedness. Moreover, this criterion may vary as a function of group. We therefore analyzed the data taking this variability into account, by examining the proportion of ‘yes’ responses, the latencies to make a related (‘yes’) decision, and the latencies to make an unrelated (‘no’) decision, as three separate dependent measures. We expect increased relatedness to lead to a higher proportion of ‘yes’ responses, shorter latencies to make a related ‘yes’ response, and longer latencies to make an unrelated ‘no’ response.

A secondary issue to be examined in the current experiment is the influence of the location of the modifiers within the pair of expressions. The modifier most often preceded the ambiguous word in both the first and the second expressions (e.g., *expiration date–dinner date*), but on occasion it followed the ambiguous word (e.g., *left foot–foot long*). The location of the modifiers within the pair of expressions could influence processing because it acts as the disambiguating context that either precedes or follows the ambiguous word. In particular, if the modifier appears before the ambiguous word it biases one of the meanings. However, if it follows the ambiguous word, the word is initially presented in a neutral context. This may lead to differences in relatedness judgments, because participants may be more likely to judge neutral words as related. Note that the role of the modifier in restricting activation may depend on the meaning dominance of the word (e.g., Duffy et al., 1988), which is beyond the scope of this investigation. Nonetheless, we can compare pairs with modifiers only appearing second, to pairs where the modifier appeared first in one or both expressions.

In addition, the location of the disambiguating context may be differentially important for polysemous versus homonymous words. Specifically, Frazier and Rayner (1990) examined

reading times for sentences including ambiguous words with multiple meanings (homonyms) or words with multiple senses (polysemous words). The disambiguating context (a clause) could precede or follow the ambiguous word. Reading times were longer when disambiguating context followed a homonym than when it preceded the homonym, but the location of the disambiguating context influenced reading times of polysemous words less than reading times of homonyms. Thus, the location of the modifiers in the current experiment may be more important for words with low intra-word sense relatedness. Note, however, that because we instantiate context using a single word, rather than a clause as in the Frazier and Rayner (1990) study, an effect for modifier location may be more difficult to detect.

To summarize, the proportion of related ‘yes’ responses, and latencies to make ‘yes’ and ‘no’ judgments will be analyzed, examining the interaction between group (English monolinguals, English-Spanish bilinguals, Spanish-English bilinguals) and condition (split vs. joint translation). The effects of relatedness ratings and that of modifier location will be examined in each of these measures.

## **3.2 METHOD**

### **3.2.1 Participants**

Participants in this experiment were those who participated in the eye-tracking experiment described in Chapter 2. These included 30 monolingual English speakers (ME), 30 English-Spanish (ES) bilinguals, and 30 Spanish-English (SE) bilinguals (see Table 1 above).

### 3.2.2 Materials

A set of 185 ambiguous English words were selected from available research on within-language ambiguity (e.g., Klepousniotou et al., 2008; Nelson, McEvoy, Walling, & Wheeler, 1980; Twilley, Dixon, Taylor, & Clark, 1994), such that some of them have a single translation in Spanish that captures both meanings/senses of the English word ('joint translation' condition, e.g., 'cuerpo' captures both the administrative and the biological meanings of the word 'body'; 84 items), and some have two Spanish translations ('split translation' condition, each encompassing one of the senses/meanings of the English word; 101 items). Translations in Spanish were determined by consulting two highly proficient Spanish-English bilinguals.

Two modifiers were selected for each ambiguous word, each highlighting a different sense or meaning of the word, for example, 'human body–administrative body'. These pairs of expressions were presented to a separate group of 20 monolingual English speakers who rated the meaning similarity of the two senses on a scale from 1 (complete different) to 7 (exactly the same) (see Appendix F). Based on these normative data the meanings of ambiguous English words vary in their semantic relatedness ( $M=2.7$ , range=1.1-5.7). These relatedness ratings were used as a predictor in the model analyzing the experimental data to ensure that differences between the 'joint translation' and 'split translation' conditions are not due to baseline differences in meaning relatedness in English. Additionally, because expressions varied in their structure and length, we used the modifiers' length and position within the expression as covariates in the model, along with the ambiguous words' length and log Kucera-Francis frequency (taken from the MRC database, Wilson, 1988). Example stimuli and expressions' characteristics are presented in Table 10; the full set of expressions is available in Appendix G.

Two versions of the stimuli were used, such that the order of the expressions within each pair was switched from one version to the next (e.g., ‘human body-administrative body’ in one version, and ‘administrative body-human body’ in the other version). Approximately half of the participants in each group completed each version (16-13 ME; 13-16 ES; 15-15 SE). Each participant saw each ambiguous word in only one pair of expressions.

**Table 10.** Experiment 2 example stimuli and characteristics by condition.

	Joint-Translation	Split-Translation
Number of items	84	101
Example	housing market – flea market	expiration date – dinner date
<i>Spanish Translation/s</i>	<i>mercado</i>	<i>cita – fecha</i>
Semantic Relatedness Rating (1-7 scale)	3.23 (1.26)a	2.23 (.92)b
Average Length (in letters)	6.29 (2.18)a	4.63 (1.21)b
Average Frequency (KF)	115.23 (124.99)a	125.24 (157.35)a
Average Concreteness	485.97 (107.15)a	481.08 (99.51)a
Average Modifier’s Length	6.76 (1.92)a	6.35 (1.68)a
Number of expressions in which the modifier followed the ambiguous word	16	39

*Note:* Means in the same row that do not share sub-scripts differ at the  $p < .05$  level.

### 3.2.3 Procedure

This task was administered in a quiet room following the sentence reading task described in Chapter 2. On each trial, participants were presented with a pair of expressions, each including an ambiguous English word along with a modifier of its meaning or sense, all at once (e.g.,

‘kitchen cabinet – presidential cabinet’). They were asked to decide if the two senses described in these expressions were related in meaning, by pressing ‘yes’ or ‘no’ as quickly and accurately as possible. ‘Yes’ responses were always made with the dominant hand. Participants were informed that this was not a test of their knowledge but rather that we were interested in their intuitions. Each trial began with a fixation cross in the center of the screen for 1500 ms followed by the pair of expressions, until a response was made or 10 seconds had elapsed. Two examples were provided in the instructions and five practice trials were presented to allow participants to become comfortable with the task. Because we were interested in participants’ subjective judgments of relatedness, feedback was never provided. One hundred and eighty-five experimental trials were presented in randomized order (E-Prime software, Psychology Software Tools, Pittsburgh, PA), interleaved with three short breaks.

As previously described in Chapter 2, following this task, participants continued the general procedure and completed proficiency tests in English and Spanish (for bilinguals) as well as working-memory and non-verbal intelligence tasks (see Appendix D).

### **3.3 RESULTS**

#### **3.3.1 Data analyses**

Data from two participants were lost; analyses were therefore performed on a final set of 88 participants (29 ME; 29 ES; 30 SE). Reaction times shorter than 200 ms were removed, constituting less than 1% of the data. Analyses were performed using linear mixed effects models as implemented in the lme4 library in R (Baayen et al., 2008; R Development Core

Team). The models included random effects for participants and items, and fixed effects for group, condition, and the interaction between them. For the effect of group, ME speakers were established as the reference, such that we examined the difference between ES bilinguals and ME speakers and the difference between SE bilinguals and ME speakers. In addition, participants' age and item characteristics (length and log Kucera-Francis frequency of the ambiguous word, length of first and second modifiers and their position within the expression) were included as covariates. Importantly, semantic similarity ratings of the expressions obtained from a different group of monolingual English speakers in a norming experiment (see Appendix F) were included in the model to account for any baseline differences between the split- and joint-translation conditions. Prior to analyses, covariates were centered and the semantic similarity ratings of the expressions were log transformed to reduce skewness.

A model including main effects was fit first, followed by a model that included the theoretically-important interaction between condition and group. The coefficient estimates from these models are reported in Table 11. To examine if changes in intra-sense relatedness are modulated by baseline differences in semantic relatedness (i.e., differ for homonyms and polysemous words), a second set of models was tested, in which relatedness was allowed to interact with condition and group. A model including two-way interactions was followed up by a model including the three-way interaction among condition, group, and relatedness.

Table 12 presents the coefficient estimates from these two models. In what follows only significant fixed effects are reported.

**Table 11.** Coefficient estimates for semantic relatedness judgments.

	'yes' Probability	'yes' RT	'no' RT
Intercept	-1.332**	8.120**	7.937**
Participants' Age	-0.019*	0.010**	0.005±
Word Length (in letters)	0.005	0.015*	0.021**
Word log KF frequency	-0.027	-0.020*	0.003
1 <sup>st</sup> Modifier's Length	-0.009	0.010**	0.012**
2 <sup>nd</sup> Modifier's Length	0.020	0.008*	0.009**
1 <sup>st</sup> Modifier's Location [after]	-0.153	0.092**	0.042*
2 <sup>nd</sup> Modifier's Location [after]	-0.224±	0.012	-0.032*
Relatedness Ratings	3.677**	-0.163**	0.166**
Group [ES]	0.211	-0.150*	-0.091
Group[SE]	0.296	0.068	0.126±
Condition [split]	-0.080	-0.011	0.012
Group [ES]: Condition [split]	-0.102	0.015	-0.001
Group [SE] : Condition [split]	0.315*	0.002	0.038*

*Note:* Coefficient for 'yes' probability were estimated with a logistic regression. For the RT analyses, coefficients were estimated on the log transformed data, and significance is based on pMCMC. ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

**Table 12.** Coefficient estimates for semantic relatedness judgments including relatedness interactions.

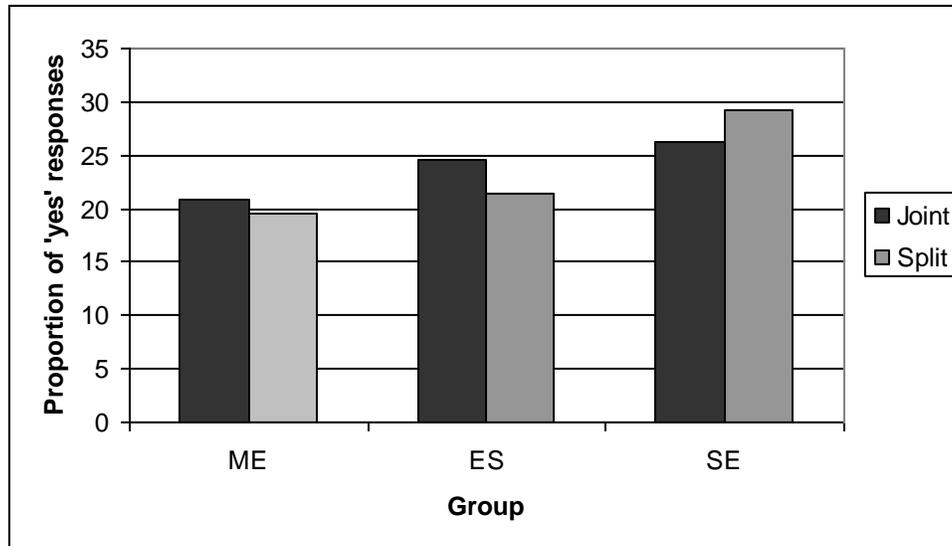
	'yes' Probability	'yes' RT	'no' RT
Intercept	-1.485**	8.137**	7.939**
Participants' Age	-0.019±	0.010**	0.005±
Word Length (in letters)	-0.022	-0.021*	0.003**
Word log KF frequency	0.004	0.015*	0.021
1 <sup>st</sup> Modifier's Length	-0.007	0.009**	0.012**
2 <sup>nd</sup> Modifier's Length	0.022	0.007*	0.009**
1 <sup>st</sup> Modifier's Location [after]	-0.155	0.093**	0.043*
2 <sup>nd</sup> Modifier's Location [after]	-0.226±	0.012	-0.032*
Relatedness Ratings (log)	4.218**	-0.220**	0.136**
Group [ES]	0.192	-0.121*	-0.090
Group[SE]	0.625*	0.064	0.127±
Condition [split]	0.038	-0.034	0.008
Group [ES]: Condition [split]	-0.088	-0.001	0.013
Group [SE] : Condition [split]	-0.004	0.011	0.038*
Condition [split] : Relatedness Ratings	-0.191	0.147**	0.027
Group [ES] : Relatedness Ratings	0.095	-0.060	0.055*
Group [SE] : Relatedness Ratings	-1.170**	0.007	0.001
Condition [split] : Group [ES] : Relatedness	-0.465	0.097	0.013
Condition [split] : Group [SE] : Relatedness	-0.471	-0.049	0.037

*Note:* Coefficient for 'yes' probability were estimated with a logistic regression. For the RT analyses, coefficients were estimated on the log transformed data, and significance is based on pMCMC. ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

### 3.3.2 Response probability

To determine if the probability of responding ‘yes’ is influenced by relatedness, group, and condition, a logistic regression model with the above predictors was fit to the data. The results show that as expected, as the relatedness of the expressions (based on ME speaker norms) increased, the probability of responding ‘yes’ increased. Further, as participants’ age increased, the probability of responding ‘yes’ decreased. In addition, when the modifier appeared after the critical word in the second expression (e.g., left foot-foot long), the probability of responding ‘yes’ marginally decreased.

Critically, the interaction between condition and group was significant, suggesting that the SE bilinguals, but not the other participants, were more likely to respond ‘yes’ for expressions in the ‘split’ condition compared to the ‘joint’ condition (see Figure 15). These findings, along with the significant positive slope of the relatedness judgments for this measure, suggest that SE bilinguals consider the two meanings of English ambiguous words to be *less* similar in meaning when both are captured by a single Spanish translation.



**Figure 15.** Proportion of 'yes' responses as a function of condition and group.

To examine if this effect was modulated by the baseline relatedness of the senses (i.e., was different for polysemous than homonymous words), relatedness was allowed to interact with group and condition. In this analysis, the effect of condition was no longer different for SE bilinguals compared to ME speakers. Instead, the effect of relatedness varied by group, such that for the SE group the probability of responding 'yes' increased with relatedness to a lesser extent than it did for the ME group (see Table 13).

**Table 13.** Percentage of 'yes' responses as a function of group and relatedness.

	ME	ES	SE
Low Relatedness	3.37	3.89	9.87
Mean Relatedness	18.47	21.53	29.74
High Relatedness	59.55	65.04	62.08

### 3.3.3 Reaction times

Significance was estimated based on 10,000 Markov Chain Monte Carlo samples of the posterior samples of the parameters (pMCMC, Baayen, 2008). Reaction times were log transformed prior to analyses. Analyses were first performed with response type (i.e., yes/no) as a predictor in the model to examine if it influenced speed of response. Because response type was a highly significant predictor of the latency data,  $F(1, 15890)=48.62$ ,  $p<.001$ , and it interacted with condition,  $F(1, 15890)=13.21$ ,  $p<.001$ , it was important to examine the effect of condition and group on ‘yes’ responses separately from ‘no’ responses.

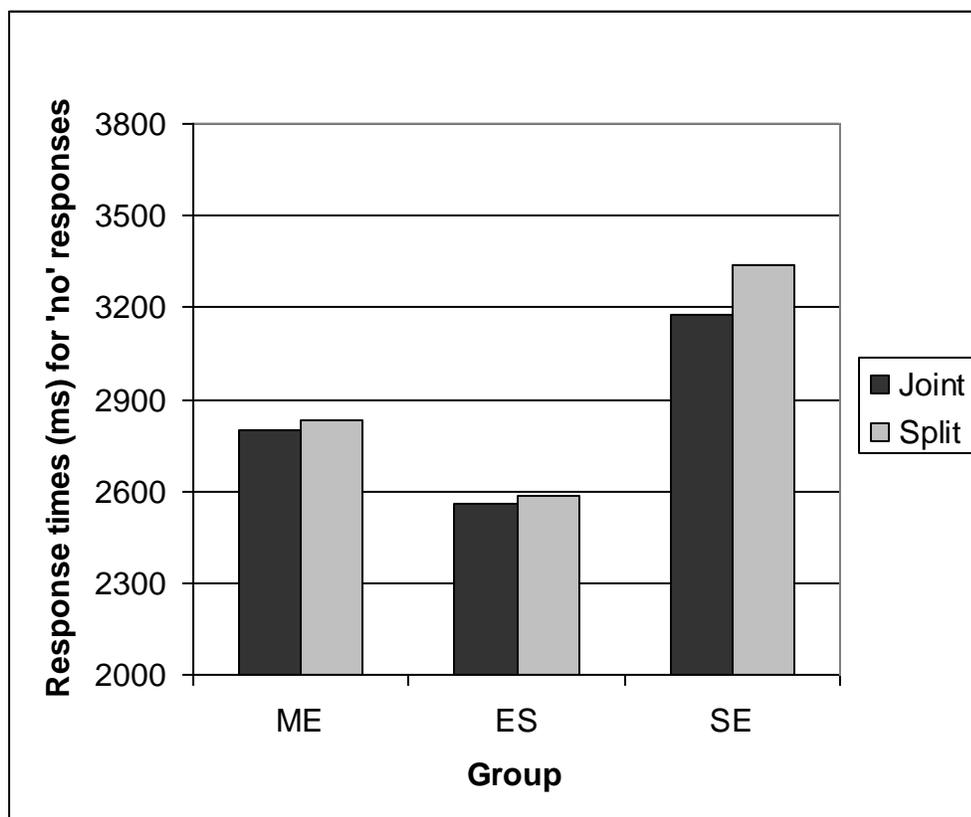
In the analyses of ‘yes’ responses, the ES bilinguals responded more quickly than the ME participants. Further, response time increased as participants’ age increased, as word and modifiers’ length increased, and as word frequency decreased. Responses were slower when the modifier followed the target in the first expression. Interestingly, as the semantic relatedness of the expressions increased, the time to make a ‘yes’ response decreased. The effect of condition and its interaction with group were not significant.

To examine if the effect of condition was different for homonyms and polysemous words, semantic relatedness was allowed to interact with condition and group. The pattern of results remained the same with the exception that the effect of semantic relatedness varied by condition, such that it was slightly stronger for joint versus split translation expressions. Critically, these effects did not interact with group.

In the analysis of ‘no’ responses, response times increased with the length of the word and its modifiers, and marginally increased with participants’ age. Furthermore, when the modifier followed the ambiguous word in the first expression, the time to make ‘no’ responses

significantly increased, but when the modifier followed the word in the second expression, the time to make ‘no’ responses significantly decreased.

Interestingly, time to make a ‘no’ response significantly increased as the relatedness of the expressions increased. SE bilinguals responded marginally more slowly than the ME speakers, but this factor significantly interacted with condition. In particular, SE bilinguals were faster to make unrelated responses for joint-translation expressions compared to split-translation expressions (see Figure 16). This finding, along with the positive slope of relatedness for the time to make a ‘no’ response, suggests that SE bilinguals consider the two meanings of an English ambiguous word to be *less* related in meaning when they share a Spanish translation, and are thus faster to indicate ‘un-relatedness’ for these joint-translation items.



**Figure 16.** Response times for ‘no’ responses as a function of condition and group.

This effect was not modulated by the baseline relatedness of the items. In particular, when the effects of condition and group were allowed to interact with that of relatedness the pattern of results remained, such that SE bilinguals were still slower at judging split-translation expressions as unrelated compared to joint-translation expressions, and this effect did not vary by relatedness of the expressions. Finally, the effect of relatedness was stronger for ES bilinguals than for ME speakers.

### 3.4 DISCUSSION

The results of the current experiment show that translation status can influence bilinguals' relatedness judgments for intra-word senses. In particular, SE bilinguals judged joint-translation expressions, which instantiate two senses of an ambiguous English word that are captured by a single Spanish translation, to be *less* related than split-translation expressions, which instantiate senses that correspond to separate Spanish words. This cross-language influence was evident in the proportion of 'yes' responses, such that joint-translation expressions were judged as related less often than split-translation expressions by SE bilinguals, in comparison to ME speakers. Furthermore, the results from the latency data for 'no' responses converge with this finding, in that joint-translation expressions were judged as 'unrelated' more quickly than split-translation expressions. This latter finding held regardless of the baseline semantic relatedness of the senses, as judged by ME speakers.

The validity of the task as reflecting participants' relatedness judgments is supported by the effect of relatedness ratings, obtained from monolingual norms, across all three dependent measures. First, pairs that were rated as more related in meaning in the norming experiment by

ME speakers were more likely to be judged as ‘related’ in this task by individuals in all three participant groups. Second, relatedness ratings were also a significant predictor of the reaction time data. In particular, higher ratings speeded responses on ‘related’ trials, and slowed responses on ‘unrelated’ trials, such that participants were faster to indicate a pair was related and slower to indicate it was unrelated when it received a higher rating in the norms. These consistent findings support the semantic nature of the task.

The main goal of the current experiment was to examine how translation status affects intra-word sense relatedness. We found reduced relatedness ratings for joint-translation senses compared to split-translation senses for SE bilinguals. This finding is consistent with the inhibition account, by which two senses sharing a label come to inhibit each other because only one is appropriate in any given context (e.g., Chwilla & Kolk, 2003; Elston-Güttler et al., 2005b). Presumably, the shared label in Spanish led to suppression between the two meanings, such that SE bilinguals considered them to be less related than two meanings that did not share a Spanish lexical form. Because in the current experiment meanings in both conditions shared an English lexical form, we cannot evaluate the effect of a shared label in English.

As mentioned in the introduction, it is more natural to assume that two unrelated meanings of homonyms inhibit each other than it is to assume that two related senses of polysemous words inhibit each other, because only the former are likely to be mutually exclusive in a vast majority of possible contexts (for discussion, see Degani et al., 2011; Klein & Murphy, 2001). Nonetheless, there is some evidence to suggest that even related senses of polysemous words may inhibit each other (Klein & Murphy, 2001; Pykkänen et al., 2006). For instance, Klein and Murphy (2001) found that sensicality judgments to polysemous words were slower and less accurate when the same word had been previously presented with a different sense than

when it had been previously presented in a neutral context (but see Klepousnioutou et al., 2008 for a different pattern). In the current experiment, the inhibitory effect for a joint (Spanish) label was not modulated by the baseline relatedness of the meanings. It is still possible, however, that a more sensitive measure would tap such relatedness modulation in the time-course of the effect, or in the mechanism underlying this inhibition. Indeed, the MEG results of Pylkkänen et al.'s (2006) study showed that inhibition between unrelated meanings was reflected by left-lateralized M350 modulation, typically associated with lexical or morphological access. Sense competition for related senses of polysemous words was presumably reflected in a (less-typical) right-lateralized M350 modulation.

An alternative explanation for the difference between joint and split translation items relies on the learning history of these items for SE bilinguals. In particular, a Spanish speaker who learns English as an L2, learns to map two words (e.g., 'anillo' and 'timbre', corresponding to diamond ring and a loud ring) to a single shared English lexical form ('ring'). Previous work has shown that two words that share a translation in a bilingual's other language become more related in meaning (e.g., Chapter 2; Degani et al., 2011; Elston-Güttler & Williams, 2008; Jiang, 2002; 2004; but see Elston-Güttler et al., 2005b). It is conceivable that this shared-translation effect is in some way an 'overshoot', such that the two meanings become more related as a function of the shared-translation than they would have had they shared a label in Spanish from the get go. That is, in the case of the joint condition, two meanings already share a label in Spanish, and no change occurs for these meanings as a function of learning an additional joint English label. Moreover, because these two senses share a label in their L1, speakers may have had the opportunity to notice the unique distinctive features of each sense and their commonalities, and thus a shared English label has little effect on the intra-word senses of these

items. Thus, split-translation meanings grow more similar in meaning as a function of a shared label in English, but joint-translation meanings do not undergo any change. As a result, SE bilinguals judge split-translation meanings to be more related than joint-translation meanings. The current data do not allow us to assess this possibility, because the use of English expressions for meaning instantiation precludes the possibility to obtain relatedness judgments/ratings from monolingual Spanish speakers. Further, there is no direct comparison in this experiment between the split-translation condition and two Spanish meanings with two (different) English translations. Future work could track relatedness judgments for bilinguals over time, or could employ a training paradigm to examine how intra-word sense relatedness changes as a function of learning.

### **3.4.1 Language proficiency**

In the current experiment, ES bilinguals did not differ from ME speakers in their processing of joint and split-translation expressions, providing no support for the influence of learning on intra-word sense relatedness, or for the possible influence of an L2 on L1 processing. Nonetheless, such effects may surface with more proficient ES bilinguals, or with those who are immersed in an L2 (Spanish) environment. The bidirectional pattern of L2 influence on L1 processing observed by Degani et al. (2011) was present for English-Hebrew bilinguals who had been immersed in their L2 (Hebrew) environment for 20 years on average. Note, however, that the ES bilinguals in the current experiment did exhibit L2 cross-language influence in a different task (see Chapter 2). It is thus possible that intra-word sense relatedness is less susceptible to change, or that the long latencies inherent to the current paradigm obscure such effects. More-proficient bilinguals may nonetheless exhibit cross-language transfer even in this paradigm. Future

investigations could sample populations that fall between these two extremes in terms of immersion experience, and could employ more sensitive measures to examine changes in intra-word sense relatedness.

As noted, the emergence of a translation status effect may depend on the relative proficiency of bilinguals in their two languages. That is, ES bilinguals who are more proficient in Spanish may show a difference in intra-word sense relatedness as a function of the translation status of the senses in Spanish. To test this possibility, we created a composite score reflecting participants' Spanish proficiency (see Appendix D), and entered this score as a predictor to the model, allowing it to interact with the effect of condition (joint vs. split translation). This analysis did not reveal any effect of condition or an interaction of condition with Spanish proficiency.

Interestingly, however, when a parallel analysis was carried out for SE bilinguals, we found that English proficiency modulated the split versus joint effect for this population. In particular, a significant interaction between condition and English proficiency emerged, such that lower English proficiency was associated with higher probability of 'yes' responses for split- compared to joint-translation items, but higher English proficiency was associated with a reversed pattern (see Figure 17). The analyses of latencies to make a related ('yes') response did not reveal any reliable effects of interest, but the results from the latency data to make an unrelated ('no') response converge with the pattern observed in the probability to make a 'yes' response. In particular, English proficiency interacted significantly with condition, such that only lower English proficiency was associated with a latency difference in the responses to joint versus split items (see Figure 18). Together, these results show that SE bilinguals of lower English proficiency show reliable increased semantic relatedness for split over joint-translation

items, but the direction of this difference tends to disappear or change with increased English proficiency.

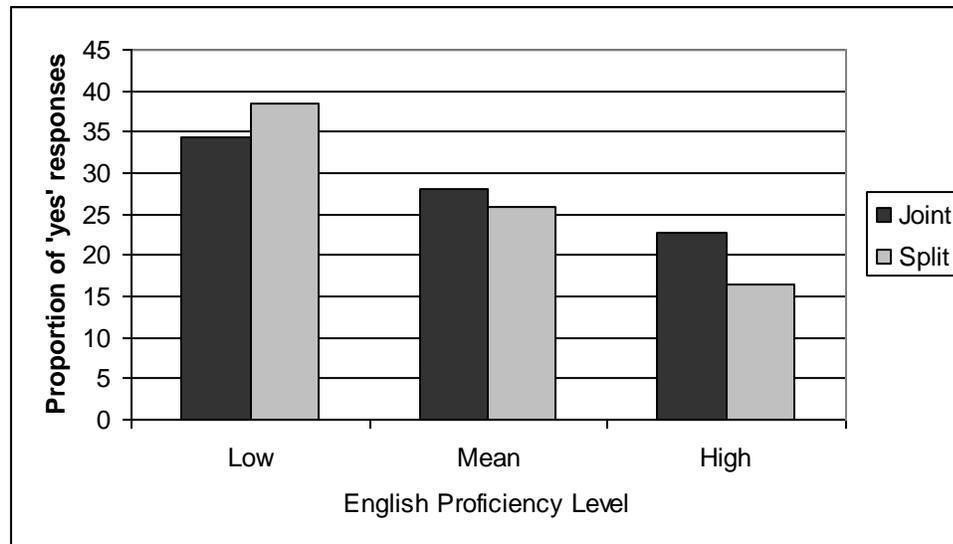


Figure 17. Proportion of 'yes' responses of SE bilinguals as a function of condition and English proficiency.

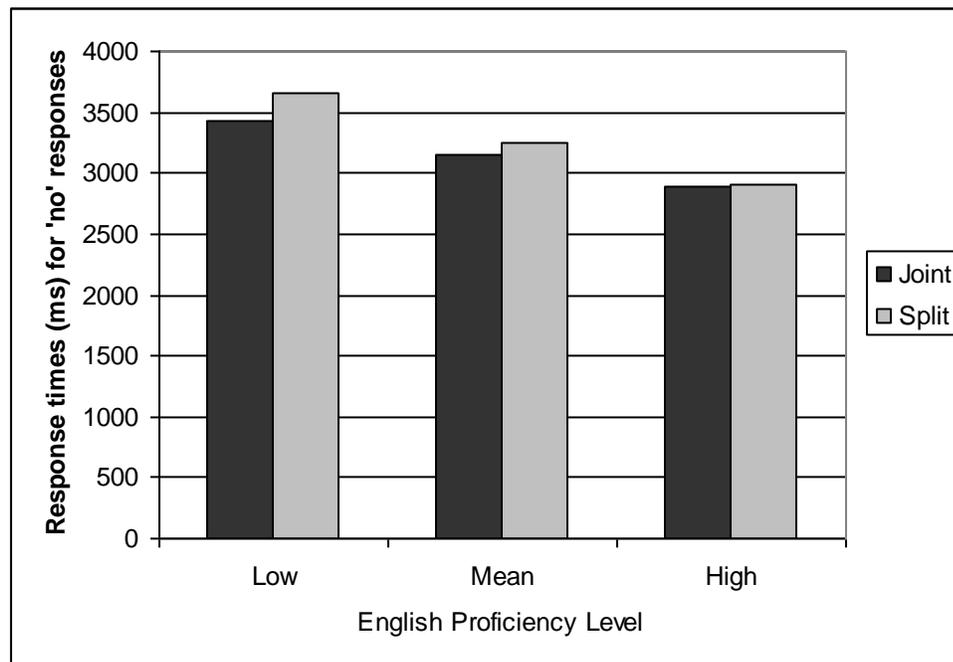


Figure 18. Response times for 'no' responses of SE bilinguals as a function of condition and English proficiency.

These findings are consistent with the results of Elston-Güttler et al. (2005b), who found a cross-linguistic effect that was observed primarily for participants of lower L2 proficiency. Specifically, native German speakers who learned English as an L2 performed a primed lexical decision task to targets preceded by a prime that shared a translation with the target in German (e.g., pine-jaw, both translated to German as Kiefer), or with an unrelated prime. Participants of lower proficiency showed reversed priming for the shared homonym condition both in and out of context, as reflected in both RT and N200 ERP modulation. The higher-proficiency participants did not show any significant ERP modulation. The authors suggested that this might be due to increased control of the higher-proficiency group, which may have decreased cross-language influence in this group.

Furthermore, Figure 18 shows that increased English proficiency was associated with faster ‘no’ responses overall in this task. Such reductions in RT may make it more difficult to identify differences between joint- and split- translation conditions. This may further explain why ES bilinguals, who are highly proficient in English, do not exhibit significant differences between conditions. Future variations in task parameters may make such differences more salient.

### **3.4.2 Location of the modifiers**

A secondary issue examined in the current experiment was the effect of modifier location on intra-word sense relatedness judgments. In particular, within each expression the modifier typically preceded the ambiguous word, but on occasion followed the ambiguous word. Because the modifier served as the disambiguating context for the word, we postulated that it might lead to reduced semantic relatedness when it preceded the word compared to when it followed it.

Following Frazier and Rayner (1990), who observed reduced effects for the location of the disambiguating context of polysemous words compared to homonyms, we further examined if this effect interacted with the baseline relatedness of the meanings.

The results showed that the effect of modifier location depended on whether it followed the ambiguous word in the first or in the second expression. Specifically, when the modifier appeared after the ambiguous word in the second expression (e.g., tsunami wave–wave goodbye) the probability of responding ‘yes’ decreased and the time to make a ‘no’ response decreased. This suggests that with no intervening words between the repetitions of the ambiguous word, the expressions were judged as less related. Conversely, when two words separated the repetition of the ambiguous word (foot long–left foot) response times were increased for both ‘yes’ and ‘no’ responses, suggesting more difficulty in decision processes. These effects did not interact with relatedness in any measure, suggesting that it was the same for homonyms and polysemous words. These effects are not entirely captured by an explanation that is based on the location of the disambiguating context. Rather, it seems that because the two expressions are presented for judgment simultaneously, the effects depend on the combination of modifier location within the first and the second expressions. Nonetheless, the results demonstrate that this factor can exert an influence on both semantic relatedness (with zero intervening words between the repeated ambiguous word), and decision processes involved in this task (with two intervening words between the repeated ambiguous word).

### **3.4.3 Baseline semantic relatedness**

In the current experiment we chose to examine intra-word sense relatedness as a continuum rather than to divide the stimuli into groups of homonyms and polysemous words. One can claim

that joint translation words are more likely to be polysemous words because rules for sense extension are often not only consistent within a language but are also shared across languages. For example, many polysemous words follow the container-content relation, where the container (e.g., ‘can’) can encompass its content, as in ‘She drank the whole can’. Therefore, two senses are likely to share a label both in English and in Spanish (i.e., joint-translation) if their relationship follows such a predicted rule. Nonetheless, other, less regular relations can also characterize polysemous words. Klepousniotou and Baum (2005) for instance, discuss and empirically compare regular metonymically polysemous words (e.g., can) to metaphorically polysemous words (like ‘lip’), for which sense extension is more irregular. Thus, some, but not all, polysemous words may share a label in another language. Furthermore, although homonyms are thought to have been accidentally created when two lexical entries share form, these ‘accidents’ could have carried over in both English and Spanish, due to their shared etymologies. Therefore the two unrelated meanings of homonyms may similarly share or not share a Spanish translation. Indeed, our norming experiment shows that both split and joint translations varied in their relatedness ratings, with a range of 1.10-5.70 for split translations and a range of 1.05-5.33 for joint translations. Critically, because relatedness ratings were entered as a predictor in the model, the effect of condition is not due to mean differences between split ( $M=2.23$ ,  $SD=0.92$ ) and joint ( $M=3.23$ ,  $SD=1.26$ ) translation items.

One potential caveat is the role of form overlap in the present results. In particular, because two unrelated meanings presumably share a label in both English and Spanish due to their shared etymologies, the translation for these unrelated senses are more likely to share form between English and Spanish (e.g., ‘operación’ in Spanish to capture the military and the mathematical senses of the word ‘operation’). Moreover, previous studies have shown that

bilinguals' semantic representations may be influenced by the form overlap across translations of different meanings of ambiguous words. Specifically, using a sentence generation task, Arêas Da Luz Fontes and Schwartz (2010) showed that Spanish-English bilinguals tended to access homonyms' meanings that share lexical form with Spanish (e.g., the weapon meaning of the word 'arm' is translated to Spanish as 'arma') more frequently than would be expected based on monolingual norms. Future studies could examine more closely whether and how form overlap between the English and Spanish translations impact the split-translation effect.

To conclude, the present experiment demonstrates an influence of translation status on intra-word sense relatedness of bilingual speakers. Joint-translation senses, which correspond to a single translation in Spanish, were less likely to be judged as related, and were judged to be unrelated more quickly, compared to split-translation senses, which map onto separate Spanish translations. This cross-language influence was present regardless of initial baseline relatedness of the senses, spanning the continuum of semantic relatedness from homonyms to polysemous words.

This influence of a bilingual's other language demonstrate the interconnectivity between multiple languages of bilinguals speakers, and supports the notion that bilinguals may differ from monolinguals in processing the meanings of words (see also e.g., Ameel et al., 2005). In the current experiment we observed only an effect of L1 on L2 representation that was stronger for bilinguals who were less proficient in L2. Whether immersion in an L2 is indeed a prerequisite for bidirectional influences to emerge in semantic relatedness of intra-word senses awaits future research.

#### 4.0 GENERAL DISCUSSION

Bilingual word meanings are evidently susceptible to bidirectional cross-language influences, highlighting the dynamic nature of the bilingual lexicon. In Experiment 1, the mappings of words to meanings in L2 clearly influenced the way in which bilingual speakers processed words in L1. The pattern observed for ES bilinguals in that experiment further shows that such influence of L2 on L1 can be present for bilinguals who learned their L2 later in life, and who are immersed in their L1 at the time of testing. These findings extend previous research (Degani et al., 2011) which demonstrated bidirectional influences for bilinguals who were immersed in their L2. The present study further extends previous research in demonstrating such bidirectional influences when words are embedded in sentence context, during natural reading. Finally, the eye tracking methodology we employed in Experiment 1 allowed a detailed characterization of how the shared-translation effect unfolds over time, and as such provided support for a comprehensive account in which both lexical inhibition and increased semantic relatedness exert an influence.

More broadly, the results from Experiment 1 emphasize the need to systematically investigate reading processes in L1 and L2. The patterns we observed clearly show that native Spanish speakers (who are proficient in English) exhibit very different reading behavior when they read in English, with longer durations and fewer skips than native English speakers (see also e.g., Keating, 2009). These SE bilinguals nonetheless demonstrated the influence of L1 on L2 processing in both eye movements and sentence naturalness ratings. Of particular interest to the

questions raised in the current study, we further observed that native English speakers who became proficient in another language, show different reading behaviors than ME speakers (see Dussias & Sagarra, 2007, for similar findings in syntactic processing). Although both of these groups were reading in their L1, ES bilinguals exhibited more skips and more regressions than ME speakers. More research is needed to examine if such effects are consistent across different populations, and identify the source of these differences.

Experiment 1 suggests that the semantic relatedness of words depends in part on the mapping of words to meaning in a bilingual's other language. Experiment 2 addressed the same question, but attempted a more direct examination of the relatedness of meanings/senses. Pairs of expressions instantiating different senses of ambiguous English words were judged for relatedness. The results indicate that the relatedness of intra-word senses is guided by the translation status of these meanings in a bilingual's two languages. Two meanings of an ambiguous English word were more related when two separate Spanish translations captured them, compared to two meanings that were captured by a single joint Spanish translation. Because this effect was insensitive to the baseline relatedness of the senses, it is less likely to be due to inhibition between the senses. Specifically, because in many contexts more than one sense of polysemous words is appropriate, the need for lexical inhibition in such cases is questionable. Instead, the results of Experiment 2 suggest that the shared-translation effect may be strong enough to create an 'overshoot' in relatedness. That is, when two meanings that are initially mapped onto separate lexical forms in the L1, are captured by a shared L2 label, they grow more related to each other. When the two meanings are initially captured by a shared label in the L1 (ambiguous word), the additional shared label in L2 may be less effective. Thus, the results of both experiments may be due to the same mechanism of increased relatedness due to a shared

label, and this effect may be stronger in bringing together two separate entries than it is in bringing together two meanings that are already captured by a shared L1 label.

The contribution of Experiment 2 is also in the novel paradigm it employs to examine intra-word sense relatedness. The coherent correlations of semantic relatedness ratings with the proportion of ‘yes’ responses and with the latencies to make ‘yes’ and ‘no’ responses validate this paradigm as reflecting participants’ semantic judgments. This paradigm may prove useful for other investigations, such as those aimed at identifying whether learning can lead to changes in intra-word sense relatedness.

Both experiments described in this thesis demonstrate cross-language influences in bilingual semantic processing, but some differences emerged between the two paradigms. Specifically, the influence of L2 on L1 processing was observed only in the eye tracking record of Experiment 1, which sampled participants’ processing over time. Sensitive measures may thus be required to capture influences on processing in the native language. Conversely, L1 influence on L2 processing was most prominent in global measures in both experiments, namely the naturalness ratings of Experiment 1 and the RT measure on ‘no’ trials of Experiment 2. This could be because differences in reading behavior in L1 and L2 made it difficult to compare eye movements of SE and ME speakers. Further, variability in L2 reading proficiency may have obscured some of the systematic effects of translation status in the eye tracking experiment.

Although the two experiments employed very different tasks and stimuli, both in fact instantiated contexts for the words in question, such that meaning selection was obligatory. In Experiment 1 this was achieved via sentence contexts and in Experiment 2 via modifiers. In doing so, the present experiments extend previous work on cross-language influences for de-contextualized words (e.g., Degani et al., 2011; Jiang, 2002; 2004) and contribute to the growing

body of studies examining such influences in more naturalistic contextualized settings (e.g., Elston-Güttler et al, 2005b; Elston-Güttler & Williams, 2008)

To conclude, the present thesis reveals several key findings. First, bilingual word meanings are subject to cross-language influences, and such influences appear to be bidirectional. That is, an L2 learned later in life may still shape the organization of semantic representations in the bilingual lexicon. Therefore, as one attempts to learn an L2, changes are likely to surface in processing of the native language. Second, these cross-language influences are present even when participants process words embedded in sentences, in a natural reading task. They are therefore not due to artificial experimental tasks and are likely present in bilinguals' daily language use. Third, the study reported here show that bilinguals may differ from monolinguals in subtle ways, such as the semantic overlap of different senses or meanings of ambiguous words. Rather than assume that what we know from monolingual research applies 'as is' to bilingual speakers, the current approach is preferred, one in which these issues are examined empirically, aiming to identify if, when, and how cross-language interactions develop in the bilingual mind.

## APPENDIX A

### STIMULI CREATION AND NORMING FOR EXPERIMENT 1

#### A.1 OVERVIEW OF NORMING EXPERIMENTS

The goal of the norming experiments described below was to identify the best set of word triplets, each including an *intended shared-translation* (ITS) word, a *shared-translation replacement* (STR) which shares a Spanish translation with the ITS, and a matched *different-translation replacement* (DTR). Further, a sentence was developed for each word triplet, such that it would be a natural-sounding sentence with the ITS completion but similarly unnatural-sounding with either an STR or DTR completion. Stimulus selection was informed by an extensive set of norms gathered from monolingual English speakers and bilinguals of Spanish and English. In particular, word pairs were first normed for semantic and form similarity by monolingual English speakers. Next, bilinguals of English and Spanish translated the words into Spanish to establish the degree to which the two words in each pair elicit a shared Spanish translation (i.e., translation overlap norms). Based on these, a set of 120 word triplets that varied in their semantic similarity were chosen.

Experimental sentences were then developed, and selected based on an additional set of norms, which included sentence predictability norms and ratings of sentence naturalness, collected from monolingual English speakers. These four norming experiments are described below.

## **A.2 SEMANTIC SIMILARITY OF WORD PAIRS (MONOLINGUAL ENGLISH)**

### **A.2.1 Participants**

One hundred and fifty five native English speakers completed the semantic similarity rating task for credit in an introductory psychology class. They were all born in the US, and were not exposed to languages other than English before age 10. All indicated currently being exposed to English at least 80% of the time. Moreover, although some had learned Spanish, none had average proficiency or average use scores greater than 4 (on scales that ranged from 1 'not proficient'/'hardly ever use' to 10 'very proficient'/'always use'). An additional 70 participants were replaced because they learned a second language before age 10 (18 participants), were exposed to English less than 20% of the time at the time of testing (17 participants), or because they rated their Spanish proficiency or use as above 4 on average (35 participants). Background information for participants from all norming experiments are presented in Table 14.

**Table 14.** Background characteristics for the participants in norming experiments.

Measure	Norming Experiment				
	Word Pair Relatedness (Experiment 1)	Translation Overlap (Experiment 1)	Sentence Predictability (Experiment 1)	Sentence Naturalness (Experiment 1)	Expression Relatedness (Experiment 2)
Number of participants	155 (61 males)	10 (6 males)	20 (11 males)	65 (42 males)	20 (11 males)
L1	English	English/Spanish	English	English	English
Age (years)	19.27 (3.43)	29.50 (8.02)	20.15 (2.89)	19.06 (0.92)	19.55 (2.11)
Age began L2 (years)	14.58 (1.40)	12.00 (2.75)	14.33 (2.57)	14.92 (1.61)	14.93 (1.21)
Time studied L2 (years)	3.49 (1.29)	11.80 (8.24)	3.51 (1.71)	3.27 (1.33)	3.17 (0.99)
L2 immersion (years)	0.00 (0.02)	2.15 (3.07)	0.01 (0.02)	0.00 (0.02)	0.00 (0.01)
L1 proficiency	9.56 (0.71)	9.80 (0.35)	9.78 (0.42)	9.61 (0.59)	9.70 (0.61)
L2 proficiency	3.13 (1.42)	7.88 (0.92)	3.96 (2.22)	3.38 (1.66)	3.02 (1.76)
L1 current use	9.78 (0.55)	6.12 (2.58)	9.77 (0.44)	9.71 (0.51)	9.73 (0.43)
L2 current use	1.42 (0.68)	5.46 (2.58)	1.98 (1.31)	1.55 (0.73)	1.52 (0.72)
Attitude toward reading	7.31 (2.27)	7.70 (2.54)	8.79 (1.23)	7.56 (2.28)	8.00 (1.89)
Rated reading amount	5.97 (1.70)	6.93 (1.56)	6.80 (1.35)	5.98 (1.43)	6.30 (1.43)

*Note.* Proficiency scores are the average of reading, writing, conversational, and speech comprehension ability ratings on a 10-point scale, on which 1 indicated the lowest level of ability. Current use scores are the average of speaking, writing, reading, listening to the radio, and watching TV ratings on a 10-point scale on which 1 indicated the lowest level of current use. Attitude toward reading reflect ratings on a 10-point scale, on which 1 reflects a ‘very negative’ attitude, and 10 reflects a ‘very positive’ attitude. Reading amount scores reflect the average reading for pleasure, work, and school on a 10-point scale, on which 1 reflects ‘none’ and 10 reflects ‘a great deal’.

### A.2.2 Stimuli

One hundred and fifty five potential ‘shared-translation’ pairs were selected from the Prior et al. (2007) norms or (mostly for the unrelated level – i.e., Spanish homonyms) were identified by two native Spanish speakers. One of the words in the shared-translation pair was selected to serve as the ITS and the other as the STR. These shared-translation pairs were predicted to vary in their semantic relatedness. Nonetheless, to ensure list variability in semantic relatedness, we included 82 filler pairs which had previously been rated as unrelated (all but two were in fact rated as less than 3 on a 1-7 scale). No pair was repeated, but 31 words were repeated across the entire set paired with different words. Sixteen participants rated these pairs in terms of semantic and form similarity.

To identify control pairs that do not share a Spanish translation (i.e., different-translation pairs, consisting of a DTR for each IST word), a set of potential words that match the STR on length in number of letters (on an item-by-item basis, with 5 exceptions), and approximate Kucera-Francis frequency and concreteness ratings (MRC database, Wilson, 1988) were selected. Part of speech was also matched; when the STR was POS ambiguous, a corresponding POS ambiguous DTR was selected (e.g., touch and play), so that both words could fit a sentence syntactically. Of these, a few potential DTRs were selected for each shared-translation pair, which were thought to be similarly related to the IST as the STR. For instance, for the pair ‘glue-tail’ (which share the translation ‘cola’ in Spanish), words that match ‘glue’ in lexical characteristics, and that are similarly related to ‘tail’ in meaning were identified (e.g., ‘goal’). Eight versions were created such that word repetition was avoided within a version (with the exception of one word that was accidentally repeated once). Each of the 8 versions included 101-123 different-translation pairs (i.e., unique IST-DTR pairings, e.g., goal-tail) and 49-76 filler

pairs, for a total of 171-186 items per version, presented in randomized order. Each pair was rated by a minimum of 8 participants ( $M=17$ , range =9-23).

### **A.2.3 Procedure**

The norming task was completed via a web-based interface. Participants completed the rating at their convenience, but were asked to complete it in one sitting and to avoid going back to previous items. Following two examples, they were asked to rate each word pair in terms of meaning similarity on a scale of 1 ('completely different') to 7 ('exactly the same'), and then rate the pair in terms of form (spelling and sound) similarity on a scale of 1 ('low similarity') to 7 ('high similarity'). This procedure was adopted to increase the likelihood that participants' meaning similarity would not be confounded by the form similarity of the pair (see also Eddington et al., 2011; Tokowicz et al., 2002). Although pairs in the various conditions were not explicitly matched on form similarity, care was taken to ensure that one word was not contained in the other word in the pair (e.g., pairs like history-story were avoided). Form similarity was also included as a covariate in the analyses of the main experiment. Following the rating task, participants completed a detailed language history questionnaire.

### **A.3 TRANSLATION OVERLAP (BILINGUALS)**

#### **A.3.1 Participants**

Ten participants volunteered their time to take part in this norming experiment (mean age 29.5,  $SD=8.02$ , 4 females). Five were native English speakers who learned Spanish as an L2 (English-Spanish) and five were native Spanish speakers who learned English as an L2 (Spanish-English). None had learned both languages together in the home environment. Half of the participants translated one version of the stimuli and the other translated the other version. Data from three additional bilinguals were discarded because they failed to provide translations for more than 15% of the words in that version. Background information is presented in Table 14 above.

#### **A.3.2 Stimuli**

Potential experimental pairs (shared-translation pairs and control pairs) were normed to establish the degree of translation overlap in Spanish. To this end, two lists were constructed by splitting each pair of English words into different lists, and eliminating repetitions. Each list included 239-240 English words, presented in a randomized order.

#### **A.3.3 Procedure:**

Participants completed the translation task via a web-based interface at their convenience. They were asked to provide the first Spanish translation they thought of for each of the English words. They were instructed to avoid using the dictionary, and were informed that this was not a test of

their knowledge but rather that we were interested in their intuitions. They were asked to complete the translation task in one sitting, in the order presented, and to refrain from going back or changing their answers. Following this translation task they completed the detailed language history questionnaire.

#### **A.4 SELECTION OF EXPERIMENTAL WORD TRIPLETS**

One hundred and twenty IST words were selected, paired with 120 STRs to create the shared-translation (critical) condition, and 120 DTRs to create the different-translation (control) condition. Table 3 summarizes the word characteristics for this final set of Appendix B items. The full set of items, along with their meaning and form similarity, is provided in.

For matching purposes, critical pairs, which share a translation, were divided into three relatedness levels: highly related (40 pairs), corresponding to pairs that received above 5.30 in the meaning similarity rating, with a range of 5.35-6.56; moderately related (40 pairs), with a range of 3.19-5.25; and unrelated (40 pairs) with a range of 1.00-3.13. As described in Chapter 2, relatedness was treated as a continuous predictor in the model.

Different-translation pairs were selected to minimize the difference in semantic similarity to the IST of the two replacement types (DTR vs. STR), with the restrictions that no word should be repeated in a single version of the experiment, and that different-translation pairs should not elicit a shared Spanish translation in the translation overlap norms. These restrictions, especially the latter, made it extremely difficult to maintain the matching across the shared- and different-translation conditions. This problem was primarily an issue for the highly-related level. Shared-translation pairs were slightly more related than their controls, which may lead to differences in

eye movements that are not due to a shared Spanish translation. Note, however, that such baseline differences between shared- and different-translation pairs should influence all participants, and not just bilinguals. Moreover, semantic similarity of the replacement word (STR or DTR) to the IST was included as a predictor in the analyses of the main experiment.

There were no significant differences between the shared- and different-translation conditions in mean form overlap, or in the replacement's length, Kucera-Francis frequency (raw and log) or concreteness (all  $F_s < 1$ ). There was, however, a significant difference in mean semantic similarity,  $F(1,239)=9.37$ ,  $MSE=2.37$ ,  $p=.002$ , with shared-translation pairs being more similar in meaning ( $M=3.98$ ) than different-translation pairs ( $M=3.37$ ). Further analyses by relatedness level reveal that this difference was significant only in the highly related level,  $F(1,79)=133.99$ ,  $MSE=.57$ ,  $p<.000$ , (all other  $F_s < 1$ , except form similarity in the moderately related level,  $p>.1$ ).

With respect to translation overlap, as mentioned earlier, shared-translation pairs were initially identified as sharing a translation according to Prior et al. (2007) norms or to two native Spanish speakers. The translation overlap norms collected for this study were mostly aimed at verifying that none of the selected different-translation pairs share a translation (note that five items were replaced after the norming and are therefore missing this translation overlap information, but according to the dictionary do not share a Spanish translation). In addition, the translation-overlap norms revealed that all but 50 items in the shared-translation condition received a shared translation at least once during the norms. Of these, 17 did elicit the same translation in Prior et al. (2007) norms. To increase power, we chose to keep all items, and noted the translation overlap results for later analyses. This was partly because in the current translation overlap norms only 5 participants provided translations for each English word (in

contrast to 30 participants in Prior et al.). Further, participants were instructed to provide only the first Spanish translation for each English word (see also, Eddington et al. 2011; Prior et al. 2007; Tokowicz et al., 2002; but see Degani et al., 2011) yielding a relatively restricted set of words that share a translation. This choice was made because this method has been previously used to predict performance (e.g., Tokowicz & Kroll, 2007), and to lessen the likelihood that participants will know what we were looking for (for discussion, see Tokowicz et al., 2002).

## **A.5 SENTENCE PREDICTABILITY (ENGLISH MONOLINGUALS)**

### **A.5.1 Participants**

Twenty native English speakers participated in this norming experiment for class credit (mean age 20.7 years,  $SD=3.24$ ; 9 females). All were born in the US, and were exposed to English at least 80% of the time at the time of testing. Nine other participants were excluded because they had indicated being relatively proficient in Spanish (above 4 on a 10-point scale; 4 participants), had been exposed to languages other than English before age 10 (3 participants), or indicated currently being exposed to English less than 80% of the time (2 participants).

### **A.5.2 Stimuli**

At least one sentence was constructed for each of the 120 critical word triplets, such that the sentence predicted the IST word, and did not predict the STR or DTR words. For example, the sentence “You have to be there exactly on .... to avoid punishment” is highly predictive of the

IST word ‘time’, but is not predictive of the STR word ‘hour’ (which shares the Spanish translation ‘hora’ with the IST ‘time’) or the DTR word ‘unit’. Sentences were constructed such that the target word never appeared as the first or last word in the sentence, and that the sentence had a total length of no more than 80 characters. Two versions were created, with 101 sentences truncated before the target word in each version. The predicted target word was never repeated within a version.

### **A.5.3 Procedure**

Participants completed the cloze task in the lab using a web-based interface. They were instructed to type a single word that could be the next word in the sentence. They then completed the detailed language history questionnaire (see Table 14 above). Fifteen of the participants first completed a relatedness judgment task (Experiment 3 norming experiment).

## **A.6 SENTENCE NATURALNESS (MONOLINGUAL ENGLISH)**

### **A.6.1 Participants**

Sixty-five native English speakers participated in this norming experiment toward class credit (mean age 19.1 years,  $SD=.92$ ; 23 females). All were born in the US, and were exposed to English at least 80% of the time at the time of testing. Thirteen other participants were excluded because they had indicated being relatively proficient in Spanish (above 4 on a 10-point scale; 2 participants), had been exposed to languages other than English before age 10 (7 participants),

failed to follow instructions (2 participants), or accidentally saw a debriefing screen before the end of the task (2 participants).

### **A.6.2 Stimuli**

The two versions that were used in the predictability norms were further subdivided into three variants each. In one variant, the sentence was completed with its IST (e.g., ‘time’), in one variant with the STR (e.g., ‘hour’), and in one variant with the DTR (e.g., ‘unit’). These were counterbalanced across lists, and intermixed with natural filler sentences such that half of the sentences in each variant were expected to be relatively natural (i.e., IST completion and filler sentences) and half were expected to be less natural (i.e., STR and DTR completions). In sum, six variants were created with 126-130 randomly ordered sentences each. Target words and sentence frames were never repeated within the same variant. Full sentences were presented, including a post-target region following the target word (which was not presented in the predictability norms).

### **A.6.3 Procedure**

Participants completed the rating task via a web-based interface at their convenience. They were instructed to rate each sentence on how natural it sounded on a scale of 1 (very natural) to 7 (very unnatural). They were asked to use the full range of the scale and to rate the sentences in the order in which they appeared. Three examples were provided. Following the rating task participants completed the detailed language history questionnaire (see Table 14 above).

## A.7 SELECTION OF EXPERIMENTAL SENTENCES

For most critical word triplets two potential sentences were developed and normed. The results from the predictability and naturalness norms informed the selection of one of these sentences. In particular, the sentence that elicited the IST word more often and the STR and DTR less often in the predictability norms was given higher priority for selection. Further, priority was given to the sentence for which the two replacements were rated similarly on naturalness. Forty-five sentences were modified slightly after the completion of the norms (16 of these in the part of the sentence preceding the target), mostly to ensure that the post-target region was not the end of the sentence, while maintaining sentence length of less than 80 characters. These changes are very unlikely to affect the naturalness rating of the sentences or its predictability, because they mostly entailed adding a word at the end of the sentence or slight changes such as changing ‘yesterday’ to ‘last week’.

Of the final set of 120 sentences some sentences unfortunately did not predict the IST word strongly enough (20 sentences had a predictability of 0%, and 24 sentences had a predictability of less than 50%). Moreover, 10 sentences elicited the STR 10-20% of the time. Twenty-eight sentences were rated as more natural with the STR completion relative to the DTR completion (as indicated by a difference of more than 1.5 units on a 7-point scale) and two had a reversed difference. Six sentences were rated as relatively unnatural (above 3.5 on a 7 point scale) when they included the IST. Predictability and naturalness were therefore included as covariates in the analyses of the main experiment. Characteristics of the full set of experimental sentences (120 items) are presented in Table 3. The sentences are available in Appendix C.

In the full set of 120 sentences, completions with the IST, STR, and DTR, match in overall sentence length,  $F < 1$ , but differ significantly in predictability,  $F(2,359)=264.28$ ,  $MSE=4.65$ ,  $p < .000$ , and naturalness ratings,  $F(2,359)=190.31$ ,  $MSE=1.11$ ,  $p < .000$ . In particular, pairwise comparisons with the Bonferroni correction for multiple comparisons show that whereas completions with the ISTs are significantly more predictable ( $M=5.59$ ) than completions with STRs ( $M=.10$ ) and DTRs ( $M=.00$ ), there is no significant difference between the two replacement types,  $p=.278$ . However, although the IST completions are significantly more natural ( $M=2.15$ ) than the STR ( $M=4.00$ ) and DTR completions ( $M=4.72$ ), the DTR completions are significantly less natural than the STR completions.

When broken down by relatedness level the same pattern of results holds, with the exception that there is no significant difference between the two replacements in naturalness ratings for the unrelated level,  $p=.734$ .

Comparisons across relatedness level reveal no significant differences among the highly-related, moderately-related, and unrelated sentences in terms of sentence length,  $F(2,119)=1.208$ ,  $MSE=114.758$ ,  $p=.302$ , target predictability,  $F < 1$ , or target naturalness,  $F < 1$ . However, Bonferroni comparisons following a significant effect for the difference in naturalness rating for the STR and DTR sentences,  $F(2,119)=7.662$ ,  $MSE=1.495$ ,  $p=.001$ , reveal that highly-related sentences had a larger difference between the two replacement types ( $M=1.29$ ) than the unrelated sentences ( $M=.23$ ) and marginally larger than the moderately-related sentences ( $M=.64$ ). These differences reflect the difficulty in finding appropriate DTRs for the highly-related level, and mirror the unintentional difference in semantic similarity to the ISTs of the two replacement types. As mentioned earlier, because perfect matching was not possible, sentence predictability and naturalness ratings were included as covariates in the analyses.

Forty natural filler sentences were taken from previous eye tracking studies examining semantic anomalies (e.g., Patson & Warren, 2010). Each of these sentences was followed by a yes/no comprehension question to encourage participants to read for meaning. The experimental sentences were slightly longer ( $M=68.34$ ) than the filler sentences ( $M=62.48$ ),  $F(1,399)=14.791$ ,  $MSE=83.847$ ,  $p<.000$ .

Three versions of the experiment were created, with 160 sentences presented in randomized order. All three versions included the same 40 filler sentences (followed by the same comprehension questions) and the same 120 sentence frames. Completion type (i.e., IST, STR, and DTR) were counterbalanced across versions with 40 sentences of each type. Half of the sentences were therefore generally natural (IST and filler sentences) and half were relatively unnatural (the two replacement types).

## APPENDIX B

### CRITICAL WORD TRIPLETS FOR EXPERIMENT 1

**Table 15.** Critical word triplets for Experiment 1. S. Rel: mean semantic-relatedness ratings for the shared-translation pair; S.F. Rel: mean form-similarity ratings for the shared-translation pair; Spanish: shared Spanish translation; S. TrO: number of times (of possible 5) that the shared Spanish translation was provided in the translation overlap norms; POS: part-of-speech for shared- and control words as instantiated in the sentence; Control: different-translation pair; C. Rel: mean semantic-relatedness ratings for the control pair; C.F.Rel: mean form-similarity ratings for the control pair.

#	Shared	S. Rel	S.F. Rel	Spanish	S. TrO	POS	Control	C. Rel	C.F. Rel
1	cut-court	1.00	4.06	Corte	0	n	range-court	1.68	1.50
2	treetop-goblet	1.00	1.75	Copa	1	n	blossom-goblet	1.11	2.39
3	pope-potato	1.13	4.00	Papa	2	n	king-potato	1.24	1.19
4	glue-tail	1.13	1.75	cola	1	n	goal-tail	1.28	2.61
5	flame-llama	1.19	3.19	llama	2	n	sugar-llama	1.43	1.77
6	sail-candle	1.13	1.88	vela	2	n	jail-candle	1.61	2.30
7	chamber-camera	1.19	3.19	cámara	2	n	portion-camera	1.39	1.87
8	rob-dock	1.06	2.81	atracar	0	v	bet-dock	1.33	1.75
9	invert-invest	1.31	5.13	invertir	3	v	insult-invest	1.67	3.76
10	wrist-doll	1.38	1.25	muñeca	3	n	twist-doll	1.39	1.61
11	bank-bench	1.33	3.33	banco	4	n	bark-bench	1.43	3.50
12	appointment-citation	1.38	1.38	cita	2	n	arrangement-citation	1.67	1.62
13	cape-end	1.56	1.69	cabo	0	n	cone-end	1.50	2.14
14	duck-leg	1.31	1.07	pata	0	n	spy-leg	1.32	1.84
15	bale-bullet	1.44	2.44	bala	0	n	base-bullet	1.71	3.14
16	peak-beak	1.25	6.00	pico	1	n	team-beak	1.52	3.43

#	Shared	S. Rel	S.F. Rel	Spanish	S. TrO	POS	Control	C. Rel	C.F. Rel
17	pile-battery	1.81	1.19	pila	1	n	kite-battery	1.79	1.58
18	room-piece	1.63	1.31	pieza	1	n	face-piece	1.94	2.94
19	plant-floor	1.63	1.33	planta	0	n	staff-floor	1.56	2.00
20	carry-charge	1.69	2.31	cargar	0	v	marry-charge	1.65	1.65
21	anger-cholera	1.75	1.67	cólera	0	n	guilt-cholera	1.33	1.38
22	career-race	1.63	2.31	carrera	3	n	budget-race	1.68	1.89
23	cure-minister	1.94	2.06	cura	0	n	loss-minister	1.80	1.50
24	crest-cockscorb	2.13	1.94	cresta	0	n	craft-cockscorb	2.15	2.95
25	take-drink	1.69	1.88	tomar	1	v	last-drink	2.14	1.35
26	tent-store	2.13	1.67	tienda	1	n	cake-store	2.57	1.67
27	clue-path	2.06	1.44	pista	0	n	plea-path	1.61	3.37
28	writing-deed	2.31	1.19	escritura	0	n	academy-deed	1.96	2.14
29	assistance-attendance	2.00	4.44	asistencia	4	n	atmosphere-attendance	2.10	3.55
30	range-saw	1.74	1.35	sierra	0	n	scene-saw	2.44	4.00
31	talent-sir	1.07	1.10	don	0	n	motive-sir	2.07	1.73
32	balloon-globe	2.33	2.15	globo	2	n	pyramid-globe	2.70	1.61
33	agitation-excitement	2.53	2.06	excitación	0	n	disbelief-excitement	3.47	2.06
34	touch-play	2.69	1.31	tocar	0	v	march-play	3.30	1.85
35	notice-news	2.69	2.81	noticia	1	n	signal-news	3.55	1.24
36	drive-manage	3.00	1.47	manejar	2	v	reach-manage	2.05	2.11
37	point-note	3.06	3.19	apuntar	0	v	state-note	3.44	2.72
38	mind-care	3.69	1.31	importar	0	v	help-care	4.72	1.83
39	attempt-intention	3.13	3.06	intento	0	n	message-intention	2.68	2.06
40	drive-conduct	3.25	1.47	conducir	1	v	trade-conduct	2.68	1.79
41	proof-test	3.63	1.25	prueba	2	n	grade-test	4.62	1.38
42	wear-dress	3.06	1.94	vestir	0	v	hang-dress	3.05	1.45
43	deceive-disappoint	3.06	3.25	decepcionar	2	v	destroy-disappoint	3.17	2.72
44	address-direction	3.50	1.94	dirección	4	n	traffic-direction	4.10	1.90
45	title-degree	3.31	1.53	título	0	n	minor-degree	4.00	1.90
46	argument-discussion	3.56	1.44	discusión	0	n	assembly-discussion	3.95	2.26
47	tongue-language	3.19	2.06	lengua	2	n	pencil-language	2.68	1.84
48	guilt-fault	3.63	4.13	culpa	2	n	anger-fault	3.56	1.61
49	point-period	3.88	3.44	punto	1	n	night-period	3.15	2.10
50	holiday-vacation	4.63	1.38	vacación	2	n	boating-vacation	3.60	1.75
51	sign-announcement	4.00	1.31	anuncio	0	n	post-announcement	4.30	1.50
52	pity-shame	3.80	1.06	lástima	0	n	envy-shame	3.32	1.64
53	dresser-closet	3.88	1.88	armario	1	n	garment-closet	3.83	2.11
54	sarcasm-irony	3.67	1.31	sarcasmo	0	n	crudity-irony	2.90	2.95

#	Shared	S. Rel	S.F. Rel	Spanish	S. TrO	POS	Control	C. Rel	C.F. Rel
55	meat-flesh	3.88	1.75	carne	3	n	bone-flesh	4.00	2.15
56	fabrication-invention	4.25	3.63	invento	0	n	inspiration-invention	4.19	4.67
57	card-letter	4.25	1.50	carta	1	n	mail-letter	5.35	2.13
58	sale-offer	4.25	1.69	oferta	0	n	loan-offer	4.05	1.84
59	drill-exercise	4.44	1.25	ejercicio	0	n	track-exercise	4.62	1.48
60	wife-woman	4.31	2.53	mujer	0	n	girl-woman	5.74	1.16
61	trust-confidence	4.38	1.33	confianza	2	n	pride-confidence	4.89	1.76
62	luck-chance	4.50	1.38	suerte	0	n	fate-chance	4.26	2.63
63	deny-negate	4.31	1.88	negar	2	v	veto-negate	5.15	1.80
64	mark-brand	4.38	2.00	marca	2	n	item-brand	4.21	1.89
65	judgment-trial	4.38	1.19	juicio	1	n	democracy-trial	4.05	1.33
66	cheer-toast	3.75	1.38	brindis	0	n	bacon-toast	3.65	1.95
67	hour-time	4.50	1.47	hora	2	n	unit-time	4.16	2.26
68	cause-reason	4.94	1.81	razón	0	n	moral-reason	4.55	2.20
69	balance-scale	4.38	1.75	balance	1	n	measure-scale	4.71	1.38
70	wood-forest	4.38	1.19	bosque	0	n	bear-forest	3.94	1.95
71	voucher-ticket	4.88	1.38	boleto	1	n	lottery-ticket	4.50	1.89
72	gather-join	4.50	1.25	juntar	2	v	bundle-join	4.35	1.59
73	ability-competence	4.75	1.50	capacidad	0	n	success-competence	3.94	2.33
74	wish-want	4.81	3.25	desear	1	v	hope-want	3.74	1.84
75	strength-force	4.94	1.53	fuerza	3	n	pressure-force	5.56	1.56
76	research-investigation	5.19	1.50	investigación	2	n	evidence-investigation	4.43	2.83
77	transfer-move	5.13	1.19	trasladar	0	v	conquest-move	3.52	1.33
78	coin-money	5.13	1.19	moneda	0	n	dime-money	5.55	1.70
79	meal-food	5.13	1.25	comida	4	n	meat-food	4.89	1.89
80	blouse-shirt	5.44	1.13	camisa	1	n	collar-shirt	4.39	1.53
81	danger-trouble	5.50	1.43	peligro	0	n	crisis-trouble	5.95	1.80
82	relief-alleviation	5.56	2.38	alivio	2	n	escape-alleviation	4.65	2.25
83	duty-obligation	5.25	1.06	obligación	1	n	rule-obligation	4.90	1.70
84	street-road	6.25	1.56	calle	1	n	ground-road	4.91	2.65
85	ceiling-roof	5.38	1.25	techo	2	n	cottage-roof	3.71	1.52
86	reject-refuse	6.00	3.63	denegar	1	v	oppose-refuse	5.13	2.87
87	boat-ship	5.38	1.44	bote	4	n	wood-ship	3.47	1.56
88	earth-world	5.50	1.38	mundo	0	n	plant-world	2.87	1.78
89	warmth-heat	5.44	1.63	calor	2	n	temper-heat	4.39	1.44
90	edge-border	5.56	1.56	borde	1	n	draw-border	2.95	1.90
91	chef-cook	5.44	2.87	cocinero	2	n	bait-cook	2.79	2.11

#	Shared	S. Rel	S.F. Rel	Spanish	S. TrO	POS	Control	C. Rel	C.F. Rel
92	finding-discovery	5.56	1.69	descubrimiento	2	n	advance-discovery	4.00	1.90
93	draw-sketch	5.50	1.19	dibujar	2	v	mold-sketch	4.00	1.35
94	watch-clock	5.35	2.10	reloj	4	n	sleep-clock	3.72	1.50
95	quarrel-fight	5.38	1.13	pelea	0	n	dispute-fight	6.06	1.78
96	talk-speak	6.56	2.19	hablar	5	v	call-speak	5.00	2.24
97	serpent-snake	5.56	3.00	serpiente	5	n	dryness-snake	2.61	1.39
98	plate-dish	5.75	1.31	plato	5	n	fruit-dish	3.11	1.88
99	balance-equilibrium	5.75	1.19	equilibrio	2	n	measure-equilibrium	3.35	2.00
100	jump-leap	5.75	1.81	saltar	2	v	dive-leap	5.00	1.81
101	award-prize	5.81	1.50	premio	3	n	panel-prize	2.26	3.26
102	fate-destiny	6.00	1.44	destino	4	n	soul-destiny	3.30	1.75
103	necessity-need	6.19	3.88	necesidad	3	n	salvation-need	3.50	1.60
104	hurt-harm	6.06	3.25	lastimar	1	v	kill-harm	5.55	2.40
105	tale-story	5.88	1.31	cuento	3	n	myth-story	5.25	2.30
106	company-business	5.94	1.50	empresa	1	n	service-business	4.67	2.83
107	pick-choose	5.88	1.44	escoger	1	v	shop-choose	4.05	2.24
108	army-military	5.88	2.44	ejército	2	n	fire-military	2.70	1.65
109	answer-response	5.93	2.56	respuesta	2	n	report-response	4.45	4.63
110	talk-chat	6.00	2.81	charlar	1	v	date-chat	3.78	3.29
111	vehicle-automobile	6.13	2.06	vehículo	1	n	highway-automobile	4.05	1.90
112	seat-chair	6.19	1.44	silla	2	n	desk-chair	3.75	1.90
113	autumn-fall	6.06	1.25	otoño	0	n	damage-fall	3.50	1.84
114	help-assist	6.19	1.31	ayudar	3	v	give-assist	4.61	2.06
115	rock-stone	6.13	1.88	piedra	2	n	clay-stone	3.95	1.67
116	swear-curse	6.25	1.81	maldición	0	n	media-curse	2.55	1.45
117	home-house	6.27	3.38	casa	4	n	room-house	4.21	2.58
118	film-movie	6.38	1.69	película	5	n	park-movie	1.78	1.48
119	rabbit-bunny	6.38	1.81	conejo	2	n	carrot-bunny	4.00	1.56
120	sofa-couch	6.50	1.50	sofá	4	n	wool-couch	2.84	2.00

## APPENDIX C

### CRITICAL EXPERIMENTAL SENTENCES EXPERIMENT 1

**Table 16.** Critical experimental sentences Experiment 1. IST: Intended shared-translation word; STR: Shared-translation replacement; DTR: Different-translation replacement ; Spanish: Shared translation for IST+STR.

#	Sentence	IST	STR	DTR	Spanish
1	The jester performed before the members of the royal court earlier today.	court	cut	range	corte
2	Once he sat down, the waiter filled the glass goblet with ice water.	goblet	treetop	blossom	copa
3	Maggie ordered a baked potato with sour cream and chives as a side dish.	potato	pope	king	papa
4	They played pin the tail on the donkey.	tail	glue	goal	cola
5	In the mountains of Peru, they spotted a lone llama wandering the path.	llama	flame	sugar	llama
6	To mask the smell of the burnt food, Alison lit a scented candle in the kitchen.	candle	sail	jail	vela
7	The photographer used a digital camera during the photo shoot.	camera	chamber	portion	cámara
8	After a day of sailing, Ron returned to the marina to dock his new boat.	dock	rob	bet	atracar
9	The broker was willing to invest in the man's business.	invest	invert	insult	invertir

#	Sentence	IST	STR	DTR	Spanish
10	The little girl dressed up her toy doll in a new outfit.	doll	wrist	twist	muñeca
11	After a long walk through the park, Sally sat down on the bench to rest a bit.	bench	bank	bark	banco
12	When Dan was driving too fast, he was given a minor citation by the police.	citation	appointment	arrangement	cita
13	When she got to the end of the book she was happy with how things turned out.	end	cape	cone	cabo
14	The runner's career was over when he broke his leg in the accident.	leg	duck	spy	pata
15	The little boy wished to be faster than a speeding bullet just like Superman.	bullet	bale	base	bala
16	The baby bird opened its beak so it could be fed.	beak	peak	team	pico
17	The TV remote didn't work because she forgot to change the battery two days ago.	battery	pile	kite	pila
18	The puzzle had a missing piece which I finally found.	piece	room	face	pieza
19	The host made sure to mop the floor before her guests came over.	floor	plant	staff	planta
20	Money is tight, so I need to know how much the masseuse will charge for an hour.	charge	carry	marry	cargar
21	Joe traveled to India to aid those who suffer from Cholera and other diseases.	cholera	anger	guilt	cólera
22	I trained hard but never expected to win the race for my team.	race	career	budget	carrera
23	The man graduated seminary, and was officially a minister of the church.	minister	cure	loss	cura
24	The rooster's red, dangly cockscomb wobbled as he walked.	cockscomb	crest	craft	cresta
25	After a long workout at the gym, Joe just wanted to drink a lot of water.	drink	take	last	tomar

#	Sentence	IST	STR	DTR	Spanish
26	He went to the grocery store to buy milk.	store	tent	cake	tienda
27	Greg wandered off the beaten path and into the woods.	path	clue	plea	pista
28	After he sold his house, he transferred the deed to the buyer.	deed	writing	academy	escritura
29	Don't be absent because the teacher takes attendance every day.	attendance	assistance	atmosphere	asistencia
30	He cut down the tree with a new saw he bought last week.	saw	range	scene	sierra
31	The military recruit had to say Yes Sir about a hundred times a day.	sir	talent	motive	don
32	She learned the location of countries by spinning the round globe in her room.	globe	balloon	pyramid	globo
33	The children could barely contain their excitement for the July 4th celebration.	excitement	agitation	disbelief	excitación
34	I love the violin, and always wanted to learn how to play it as a professional.	play	touch	march	tocar
35	Every night I watch the ten o'clock news on Fox and not ABC.	news	notice	signal	noticia
36	He was so overloaded, that it was not clear he could manage the extra stress.	manage	drive	reach	manejar
37	To prepare for an emergency, I mentally note the possible scenarios.	note	point	state	apuntar
38	It's sad that the situation is so bad and people don't care much at all.	care	mind	help	importar
39	I'm sorry my comment offended you, but it was never my intention to hurt you.	intention	attempt	message	intento
40	An insulated copper wire will efficiently conduct an electric current.	conduct	drive	trade	conducir

#	Sentence	IST	STR	DTR	Spanish
41	Adam felt ready for his driver's test because he practiced parallel parking.	test	proof	grade	prueba
42	For Halloween, he was planning to dress as Count Dracula.	dress	wear	hang	vestir
43	Joe was worried that his low grade might disappoint his mom and dad.	disappoint	deceive	destroy	decepcionar
44	The lost lady asked if she were going in the right direction or not at all.	direction	address	traffic	dirección
45	I went to college to earn a Bachelor's degree in math and physics.	degree	title	minor	título
46	The convention included a panel discussion of the topics.	discussion	argument	assembly	discusión
47	I only speak one language currently, though I want to learn more.	language	tongue	pencil	lengua
48	The car accident was not my fault this time, but it was the last time.	fault	guilt	anger	culpa
49	During the ice hockey game, the third period was the most intense.	period	point	night	punto
50	The students waited for the end of school and summer vacation for a long time.	vacation	holiday	boating	vacación
51	At our next meeting, I wanted to make a brief announcement about the picnic.	announcement	sign	post	anuncio
52	The politician in the scandal brought shame to his family.	shame	pity	envy	lástima
53	Kristin hung her new dress in the walk-in closet in her bedroom.	closet	dresser	garment	armario
54	Rain on your wedding day is considered irony for a good reason.	irony	sarcasm	crudity	sarcasmo
55	The doctor said that the cut was a minor flesh wound	flesh	meat	bone	carne

#	Sentence	IST	STR	DTR	Spanish
	and would heal quickly.				
56	The telephone was Alexander Graham Bell's most famous invention of his life.	invention	fabrication	inspiration	invento
57	She checked her mailbox everyday for his letter but it never came.	letter	card	mail	carta
58	The man who wanted to buy the house made a generous offer to the sellers.	offer	sale	loan	oferta
59	The two most important things to staying healthy are exercise and a good diet.	exercise	drill	track	ejercicio
60	There were twelve men and one woman in the classroom.	woman	wife	girl	mujer
61	The economic crisis led people to lose confidence in what the government does.	confidence	trust	pride	confianza
62	Everybody who messes up once deserves a second chance in the future.	chance	luck	fate	suerte
63	Our findings do not necessarily negate the theory but they weaken it.	negate	deny	veto	negar
64	Levi's is a famous brand of jeans and clothing.	brand	mark	item	marca
65	After the attempted robbery, the thief was put on trial in the city court.	trial	judgment	democracy	juicio
66	The groomsman raised a toast for the young couple.	toast	cheer	bacon	brindis
67	You have to be there exactly on time to see the ceremony.	time	hour	unit	hora
68	Cara said that she would stay, if we gave her one good reason not to leave.	reason	cause	moral	razón
69	Brad weighed himself on the bathroom scale every morning.	scale	balance	measure	balance
70	Logging is a huge problem in the Amazon rain forest in Brazil and Peru.	forest	wood	bear	bosque

#	Sentence	IST	STR	DTR	Spanish
71	The train conductor checked my ticket after I boarded the train.	ticket	voucher	lottery	boleto
72	Andy found the volunteer club interesting and wanted to join to be a member.	join	gather	bundle	juntar
73	Julia knew that her high level of competence in the field would get her the job.	competence	ability	success	capacidad
74	My favorite Christmas song is "All I want for Christmas is my 2 front teeth."	want	wish	hope	desear
75	When the apple fell, Newton concluded a gravitational force was the cause.	force	strength	pressure	fuerza
76	After the murder, the detective led an in-depth investigation to solve the case.	investigation	research	evidence	investigación
77	I knew I broke my leg because I could barely move it side to side.	move	transfer	conquest	trasladar
78	In deep debt with her friends and bank, she owed a lot of money at the moment.	money	coin	dime	moneda
79	Burger King is my favorite fast food restaurant in the world.	food	meal	meat	comida
80	Steve was clothes shopping to find a new button down shirt to wear on his date.	shirt	blouse	collar	camisa
81	After doing something wrong he knows he is in trouble and he runs.	trouble	danger	crisis	peligro
82	The pills Maxine took after her surgery helped with the alleviation of the pain.	alleviation	relief	escape	alivio
83	We don't have a moral obligation to do what we can do.	obligation	duty	rule	obligación
84	When driving, please keep your eyes on the road in front of you.	road	street	ground	calle
85	The couple climbed up to the roof to look at the stars.	roof	ceiling	cottage	techo

#	Sentence	IST	STR	DTR	Spanish
86	Despite my continuous attempts, you refuse to try the soup.	refuse	reject	oppose	denegar
87	The captain of the pirate ship ordered the captive to walk the plank.	ship	boat	wood	bote
88	The conference had scientists from all over the world present their work.	world	earth	plant	mundo
89	The house is freezing because they haven't turned on the heat yet this winter.	heat	warmth	temper	calor
90	Without a passport, they won't let you across the border into Canada or Mexico.	border	edge	draw	borde
91	The diner hired a new short-order cook for the morning shift.	cook	chef	bait	cocinero
92	Rachel loved to watch nature shows on the Discovery Channel with her family.	discovery	finding	advance	descubrimiento
93	Before you paint the final product, you should sketch it on a piece of paper.	sketch	draw	mold	dibujar
94	The ticking hands of the grandfather clock were the only sound in the house.	clock	watch	sleep	reloj
95	They had to call the police when the bar fight got out of control.	fight	quarrel	dispute	pelea
96	At the conference, the renowned biologist was asked to speak to the audience.	speak	talk	call	hablar
97	His mom was terrified, but he thought finding a garden snake was very cool.	snake	serpent	dryness	serpiente
98	Eric repositioned his satellite dish to improve reception.	dish	plate	fruit	plato
99	It takes time for a system to reach a state of equilibrium in cold temperatures.	equilibrium	balance	measure	equilibrio
100	The frog jumped onto the leaf in one giant leap from	leap	jump	dive	saltar

#	Sentence	IST	STR	DTR	Spanish
	the riverbank.				
101	Melanie won a thousand dollars as the first prize in the competition.	prize	award	panel	premio
102	Sadly children's socioeconomic status determines their destiny in the future.	destiny	fate	soul	destino
103	After the storm, our yard is in desperate need of cleaning and repair.	need	necessity	salvation	necesidad
104	When Sawyer stepped on the ant, he meant no harm to the little creature.	harm	hurt	kill	lastimar
105	"A Study in Scarlet" is a short story about the adventures of Sherlock Holmes.	story	tale	myth	cuento
106	He had a legitimate business of selling used furniture.	business	company	service	empresa
107	Given more than one solution to a problem, you must choose the best one.	choose	pick	shop	escoger
108	The Air Force is one of five branches in the American military as is the Navy.	military	army	fire	ejército
109	An increased heart rate is an autonomic response in a stressful situation.	response	answer	report	respuesta
110	Jill called me last night to chit chat about her wedding plans.	chat	talk	date	charlar
111	Mary called AAA, the American Automobile Association, when her car broke down.	automobile	vehicle	highway	vehículo
112	They put the baby on the high chair before they started eating.	chair	seat	desk	silla
113	I love the summer, but the colorful leaves in the fall are even better.	fall	autumn	damage	otoño
114	The use of cultural brokers can assist health and social work professionals.	assist	help	give	ayudar
115	Laura got a ring with her birth stone for her birthday.	stone	rock	clay	piedra

#	Sentence	IST	STR	DTR	Spanish
116	Harry the wizard overthrew his enemy by setting a magical curse on his castle.	curse	swear	media	maldición
117	The president of the United States lives in the White House in Washington DC.	house	home	room	casa
118	I like to buy popcorn when I go to the movie theater with my friends.	movie	film	park	película
119	His favorite cartoon characters are Bugs Bunny and Tweety Bird.	bunny	rabbit	carrot	conejo
120	Joe's constant laziness led his friends to call him a couch potato to his face.	couch	sofa	wool	sofá

## **APPENDIX D**

### **METHODS AND RESULTS FOR PROFICIENCY AND INDIVIDUAL DIFFERENCES TASKS**

#### **D.1 METHOD**

##### **D.1.1 Lexical decision**

This task was intended to provide a measure of participants' word recognition abilities in English and in Spanish. In this task, two letter strings were presented briefly one after the other, and participants were instructed to read both letter strings but to respond only to the second. The first string was always a real word, and participants were to press the 'yes' key if the second letter string was a real word (in English for the English version of the task, or in Spanish for the Spanish version), and the 'no' key if it was not. On each trial, a fixation cross appeared for 200 ms, followed by the first letter string for 200 ms. A blank screen was then briefly presented (200 ms), followed by the target string until participants made a response or 3 seconds had passed. An ISI of 800 ms was used. There were 120 trials presented in a random order by the computer program (E-Prime), half of which required a 'yes' response (always made with the dominant

hand). Preceding these experimental trials, 14 practice trials were presented, during which the experimenter stayed in the room to answer any questions.

English words were 6.11 letters in length on average (range 3-11), with an average (KF) frequency of 66.47 ( $SD=80.22$ ) (with a maximum of 69971), and an average concreteness rating of 481.92 ( $SD=116.65$ ) on a 100-700 scale. Nonwords for the English version were constructed to be orthographically similar to real English words, matched in length to the real words (Armstrong & Plaut, 2008). Care was taken to ensure that they were not real Spanish words.

In the Spanish version of the task, Spanish words were slightly longer ( $M=7.06$ , range 3-11) than those used in the English version, with average frequency (from LEXESP Sebastián-Gallés, Martí, Cuetos, & Carreiras, 2000; using B-pal, Davis & Perea, 2005) of 18.95 per million ( $SD=38.91$ ) and concreteness rating of 4.87 ( $SD=2.1$ ) on a scale from 1 to 7. Nonwords were matched on length and were orthographically similar to real Spanish words, but were not real English words.

### **D.1.2 Picture naming**

This task was adopted from Tokowicz (1997) in which it was used in both English and Spanish to assess L2 proficiency of English-Spanish and Spanish-English bilinguals. In this task, participants were asked to name 30 line drawings (taken from Snodgrass & Vanderwart, 1980) in one language only (e.g., in English). On each trial, a fixation cross was presented at the center of the screen until the participant initiated the beginning of the trial with a button press. At that point, the picture was presented until the participant named the picture out loud or until 2500 ms had passed. Participants were instructed to name the bare noun, and were provided with 10 practice trials. Stimuli were presented in a random order. As detailed in Tokowicz (1997),

stimuli for this task were selected to be of similar difficulty in English and in Spanish (with similar name agreement, length, and familiarity). Words varied greatly in their English frequency to allow good sampling of participants' proficiency. Bilingual participants performed the task twice, first in their L2 and then in their L1. ME speakers performed the task only in English.

### **D.1.3 Raven's progressive matrices**

This task was used to provide a measure of participants' non-verbal intelligence. Because it is believed to not rely on language abilities, it serves as a good measure on which to compare participants of different language backgrounds. Participants were presented with a computerized abbreviated version of the task (Ravens, 1960; see, e.g., Landi, 2010; Nelson, 2010, for the abbreviated version). It included 18 items, presented in order of difficulty, and was to be completed in 15 minutes. One example was presented first. On each trial, a 3 by 3 array of patterns was presented, which became increasingly complex from left to right and from top to bottom. Participants were asked to select (among 8 alternatives) the missing pattern for the bottom right corner that would complete the series.

### **D.1.4 Operation-word working memory span**

This task was used to measure participants' working memory span (Turner & Engle, 1989; see also Michael et al., 2011; Tokowicz et al., 2004) by requiring participants to remember lists of words while simultaneously judging the accuracy of solutions to simple mathematical operations. On each trial, a fixation cross (+) was presented for 1000 ms, followed by a mathematical expression, presented for 2500 ms. A question mark probe was then presented until participants

indicated whether the expression was correct (by pressing a ‘yes’ key with their dominant hand) or incorrect (by pressing a ‘no’ key with their other hand). Upon their response an L1 word was presented for 1250 ms, followed by the next trial. After the last word in the set was presented, a “RECALL” prompt was displayed briefly, at which point participants were to type in the words they remembered from that set in the order in which they appeared. Typing was not timed, and participants were instructed to omit any special characters (such as accent marks). Operation-word sets were presented in sets increasing in size from 2 to 6, with 3 sets of each size. Two practice sets (one of 4 and one of 6) preceded the experimental sets.

Because the task entails memory of words, a Spanish version and an English version were used, such that all participants completed the task in L1 (Tokowicz et al., 2004). This was done to ensure that any emerging group differences are not due to SE bilinguals’ difficulty with processing L2 words. Words in the two sets match on length, and were translations of each other.

#### **D.1.5 Language history questionnaire**

Participants completed a detailed language history questionnaire via a web-based interface. The questionnaire was adapted from the LEAP-Q measure ((Marian, Blumenfeld, & Kaushanskaya, 2007), and is available in Appendix I. Participants were asked about L1 and L2 proficiency and use, as well as questions regarding reading and computer skills.

#### **D.1.6 Vocabulary post-test**

This paper and pencil questionnaire was designed to verify participants’ familiarity with the critical English (IST and STR) words and their shared Spanish translation. This is because any

shared-translation influence is only expected if participants are familiar with the words in question and are aware these are translations of each other (see also Elston-Güttler, 2000). In this task participants were presented with a list of Spanish-English translations, and were asked to mark an 'X' next to any pair of translations with which they were not familiar (either with one of the words individually or with the words as translations of each other). Each critical Spanish word was presented twice, once with each of its English translations. Two-hundred and forty pairs were presented in a semi-randomized order such that a given Spanish word was not presented twice in a row. No time limit was imposed. The full questionnaire is available in Appendix J

## **D.2 RESULTS**

### **D.2.1 English lexical decision**

For each participant we computed an accuracy and a latency score from their performance on the English primed lexical decision task. Mean reaction time was computed for correct trials only, and after removal of trials with RTs shorter than 300 ms or longer than 2500 ms (resulting in the removal of 0.33% of correct trials for the ME group, 0.37% for the ES group, and 0.58% for the SE group).

Accuracy was high overall ( $M=.94$ ,  $SD=.058$ ). A oneway ANOVA was performed with participants as a random variable. As expected, results reveal a significant effect of group on both the accuracy data,  $F(2,89)=8.21$ ,  $MSE=.003$ ,  $p=.001$ , and the latency data,  $F(2,89)=15.40$ ,  $MSE=17529.51$ ,  $p<.000$ . Bonferroni corrections reveal that the SE group was significantly less

accurate ( $M=.91$ ) than the ME group ( $M=.94$ ), but that the ES group did not differ from either other group ( $M=.97$ ). Further, the SE group was significantly slower ( $M=895.94$ ) than both the ME group ( $M=785.06$ ) and the ES group ( $M=707.20$ ), with a marginal difference between these two native English speaker groups.

### **D.2.2 Spanish lexical decision**

For each bilingual participant we computed an accuracy and a latency score from their performance on the Spanish primed lexical decision task. Mean reaction time was computed for correct trials only, and after removal of trials with RTs shorter than 300 ms or longer than 2500 ms (resulting in the removal of 1.2% of correct trials for ES group, and 0.32% for the SE group).

Accuracy was relatively high overall ( $M=.88$ ,  $SD=.10$ ). A oneway ANOVA was performed with participants as a random variable. As expected, results reveal a significant effect of group on both the accuracy data,  $F(1,59)=55.45$ ,  $MSE=.005$ ,  $p<.000$ , and the latency data,  $F(1,59)=20.77$ ,  $MSE=30660.23$ ,  $p<.000$ . The SE group was both more accurate ( $M=.95$ ) and faster ( $M=812.26$ ) than the ES group ( $M=.81$ ;  $M=1018.32$ ).

### **D.2.3 English picture naming**

For each participant we computed an accuracy and a latency score from their performance on the English picture naming task. Mean reaction time was computed for correct trials only, and after removal of trials with RTs shorter than 300 ms or longer than 2000 ms (resulting in the removal of 7.9% of correct trials for the ME group, 9.6% for the ES group, and 21.6% for the SE group).

Note that for most of these removed trials reaction time was recorded as zero due to voice-key errors.

Accuracy was high overall ( $M=.90$ ,  $SD=.10$ ). A oneway ANOVA was performed with participants as a random variable. As expected, results reveal a significant effect of group on both the accuracy data,  $F(2,89)=22.05$ ,  $MSE=.007$ ,  $p<.000$ , and the latency data,  $F(2,89)=12.67$ ,  $MSE=18128.95$ ,  $p<.000$ . Bonferroni corrections reveal that the SE group was significantly less accurate ( $M=.82$ ) than the ME group ( $M=.93$ ) and the ES group ( $M=.96$ ) which did not differ from each other. Similarly, the SE group, was significantly slower ( $M=1044.75$ ) than both the ME group ( $M=909.80$ ) and the ES group ( $M=880.83$ ), which again did not differ from each other.

#### **D.2.4 Spanish picture naming**

For each bilingual participant we computed an accuracy and a latency score from their performance on the Spanish picture naming task. Mean reaction time was computed for correct trials only, and after removal of trials with RTs shorter than 300 ms or longer than 2000 ms (resulting in the removal of 12.2% of correct trials for the ES group, and 5.6% for the SE group). Note that for most of these removed trials reaction time was recorded as zero due to voice-key errors.

Accuracy was relatively low overall ( $M=.65$ ,  $SD=.23$ ). A oneway ANOVA was performed with participants as a random variable. As expected, results reveal a significant effect of group on both the accuracy data,  $F(1,59)=86.40$ ,  $MSE=.022$ ,  $p<.000$ , and the latency data,  $F(1,59)=65.20$ ,  $MSE=31933.70$ ,  $p<.000$ . The ES group was less accurate ( $M=.48$ ) and slower ( $M=1350.75$ ) than the SE group ( $M=83.22$  and  $M=975.00$ ).

### **D.2.5 Operation-word working memory span**

Data for one ME participant were not available for this task. In addition, following recommendations of Conway, Kane, Bunting, Hambrick, Wilhelm, and Engle (2005), only data from participants who accurately responded to the operations with accuracy of at least 85% were included in the analyses. Analyses were therefore performed on data from 79 participants (26 ME, 27 ES, and 26 SE).

We computed a weighted working memory span measure (PCU, Conway et al., 2005) as the mean proportion of words recalled in each set. Analyses of variance revealed that the three groups differed in their working memory span,  $F(2,78)=4.00$ ,  $MSE=.007$ ,  $p=.022$ . Specifically, SE bilinguals recalled fewer words ( $M=.87$ ) than the ME group ( $M=.93$ ), and marginally fewer words than the ES group ( $M=.93$ ). The ME and ES groups did not differ from each other.

These group differences in working memory should be interpreted with caution because the groups also differ in age and education. Further, based on participants' verbal comments during the task, although words were always presented in L1, the SE group had less experience in performing mathematical operations depicted via symbols such as “\*” and “/”. This might have made the task slightly more difficult for these non-native English speakers.

### **D.2.6 Raven's progressive matrices**

Data for 10 participants were not available for this task due to technical failures and time constraints (for ME speakers who completed the experiment for class credit). Analyses were therefore performed on data from 26 participants in the ME group, 26 in the ES group, and 28 in the SE group.

Results show that 6.61 sets were correctly completed on average ( $SD=3.51$ , range 0-16). Moreover, analyses of variance showed that there were no significant differences among the groups in the number of sets correctly completed,  $F<1$ .

### **D.2.7 Language history questionnaire**

Data from the language history questionnaire are available in Table 1 (see Chapter 2). Analyses of variance reveal a significant effect of age,  $F(2,89)=20.894$ ,  $MSE=72.383$ ,  $p<.001$ , such that SE bilinguals ( $M=33.63$ ) were significantly older than ES bilinguals ( $M=23.83$ ) and ME speakers ( $M=19.83$ ), which did not differ from each other. There were no significant differences among the groups in L1 self-rated proficiency,  $F(2,86)=1.179$ ,  $MSE=.186$ ,  $p=.313$ , but differences emerged in rated L1 use,  $F(2,86)=71.285$ ,  $MSE=2.285$ ,  $p<.001$ . Post-hoc tests with Bonferroni corrections revealed that SE bilinguals rated their L1 use as significantly lower than ES bilinguals ( $M=5.19$  vs.  $M=8.61$ ) and ME speakers ( $M=9.75$ ), which also differed significantly from each other.

In addition participants' attitude toward reading marginally differed by group,  $F(2, 89)=3.003$ ,  $MSE=2.845$ ,  $p=.055$ . Tests with the Bonferroni corrections revealed that SE bilinguals rated their attitudes toward reading as significantly more positive ( $M=9.30$ ) than ME speakers ( $M=8.23$ ). ES bilinguals ( $M=8.73$ ) did not differ from either other group. A similar pattern emerged in participants' rated average amount of reading for pleasure, work, and school,  $F(2,89)=4.651$ ,  $MSE=2.395$ ,  $p=.012$ , with SE bilinguals reading more ( $M=7.54$ ) than ME speakers ( $M=6.35$ ), and ES bilinguals not differing from either other group ( $M=7.18$ ).

Comparisons between the two bilingual groups reveal no significant differences in the age at which L2 learning began or L2 proficiency,  $F_s<1$ . Significant differences did emerge in

number of years of L2 learning,  $F(1,59)=6.107$ ,  $MSE=40.291$ ,  $p=0.016$ , such that SE bilinguals studied their L2 for significantly longer ( $M=13.20$ ) than ES bilinguals ( $M=9.15$ ). Further, SE bilinguals spent significantly more time immersed in their L2 ( $M=6.98$  years) compared to ES bilinguals ( $M=.33$  years),  $F(1,56)=17.938$ ,  $MSE=35.105$ ,  $p<.001$ . SE bilinguals also rated their current use of L2 as significantly higher ( $M=7.89$ ) than ES bilinguals ( $M=4.03$ ),  $F(1,56)=93.262$ ,  $MSE=2.274$ ,  $p<.001$ .

### **D.2.8 Vocabulary post-test**

For each bilingual participant we counted the number of items that were marked as unfamiliar in the vocabulary post-test. Because each Spanish word was presented twice, once with each of its English translations, there were 240 items. Overall familiarity was high, with an average of 20 ( $SD=16.36$ ) items being marked as unfamiliar. Of interest, ES and SE bilinguals differed in their performance,  $F(1,59)=18.031$ ,  $MSE=207.679$ ,  $p<.000$ , such that ES bilinguals marked more items ( $M=27.90$ ,  $SD=17.02$ ) as unfamiliar compared to SE bilinguals ( $M=12.10$ ,  $SD=11.22$ ).

### **D.2.9 Comparing groups of native English speakers**

Overall the ES bilinguals performed slightly better than the ME speakers in the linguistic tasks (lexical decision, picture naming, and accuracy on filler comprehension questions). These differences may be traced back to motivational differences between the participants performing the experiment for class credit (ME speakers) and those participating for payment (ES bilinguals). Alternatively, age and education differences between these native English speaker groups may underlie the differences in linguistic performance. Note that it is less likely that these

differences are due to language background differences (i.e., bilingualism vs. monolingualism) per se, because such differences would have been more likely in an opposite direction, with an advantage for monolinguals over bilinguals (e.g., in picture naming, Gollan, Montoya, Fennema-Notestine & Morris, 2005). This is especially true because the bilinguals performed the picture naming and lexical decision tasks after performing the same tasks in Spanish, thus having to switch back to their L1.

The numerical difference between the ME and the ES participants in the working memory task, reflecting a slight advantage for the ME speakers, could be explained based on age differences. Alternatively, the fact that this task was the first English task to be performed by the ES bilinguals after two Spanish tasks, may explain their slight disadvantage, because switching back into the L1 is known to be a difficult task (e.g., Costa & Santesteban, 2004).

#### **D.2.10 Principal component analyses of individual-difference measures**

Several individual difference measures were used in the current study. For bilinguals, these include performance (accuracy and RT) on the English and Spanish lexical-decision tasks; picture naming tasks; operation-word span task; Raven's progressive matrices task; as well as self-rated proficiency and use of English and Spanish. Scores on all measures were available for 46 participants, 22 of whom were ES bilinguals, and the remaining 24 were SE bilinguals. Because these factors correlated strongly, a principal component analysis (PCA) was applied to the data to reduce collinearity in the predictors. Indeed, Bartlett's Test of Sphericity was  $p < .001$ , indicating that the predictors were highly correlated, and suggesting that a factor analysis is warranted for this data set. Factors with Eigenvalues over 1 were extracted, and a Varimax rotation was applied to increase interpretability of the factors by increasing the likelihood that

each original test would correlate highly with only one factor. This resulted in the extraction of four orthogonal factors which cumulatively capture 73.57% of the variance in these predictors.

The rotated component matrix is available in Table 17.

**Table 17.** Rotated component matrix.

	Spanish Proficiency	English Proficiency & Use	Spanish Use	Nonlinguistic abilities
Spanish Lexical Decision RT	-0.52*	0.16	0.58*	-0.01
Spanish Lexical Decision Accuracy	0.87*	-0.24	0.09	-0.12
Spanish Picture Naming RT	-0.80*	0.28	0.16	-0.11
Spanish Picture Naming Accuracy	0.86*	-0.29	0.08	-0.09
Rated Spanish Proficiency	0.81*	-0.25	-0.32	-0.07
Rated Current Spanish Use	0.10	-0.06	-0.81*	-0.09
English Lexical Decision RT	0.37	-0.51*	0.47	-0.23
English Lexical Decision Accuracy	-0.07	0.72*	0.32	0.20
English Picture Naming RT	0.15	-0.80*	-0.02	-0.12
English Picture Naming Accuracy	-0.36	0.81*	0.08	-0.08
Rated English Proficiency	-0.38	0.84*	0.01	0.10
Rated Current English Use	-0.23	0.65*	-0.06	-0.28
Operation-Word Span (PCU)	-0.25	0.00	-0.10	0.83*
Ravens' – Number of Correct Responses	0.23	0.11	0.46	0.68*

Tests were assigned to a factor when their correlation with it exceeded .50 (see e.g., Nelson, 2010). The first factor captures participants' Spanish proficiency (but not use) with Spanish lexical decision and picture naming (both RT and accuracy) and self-rated Spanish proficiency loading highly on this factor. It accounts for 41.27% of the variance in the data. The

second factor captures participants' English abilities with English lexical decision (RT and accuracy), English picture naming (RT and accuracy) and self-rated English proficiency and use loading on this factor. It accounts for 12.45% of the variance. The third factor captures participants' Spanish use and to a moderate degree RT on the Spanish lexical decision task, accounting for 10.99% of the variance. Lastly, the fourth factor captured participants' non-linguistic abilities (working memory span and non-verbal intelligence as measured by Ospan PCU and Ravens respectively), reflecting 8.86% of the variance.

## APPENDIX E

### ANALYSES OF THE PRE-TARGET REGION EXPERIMENT 1

Coefficient estimates for the model predicting eye movement measures in the pre-target region are presented in Table 18. These are based on the highest-order model with significant or marginally significant effects.

**Table 18.** Coefficient estimates for the pre-target region.

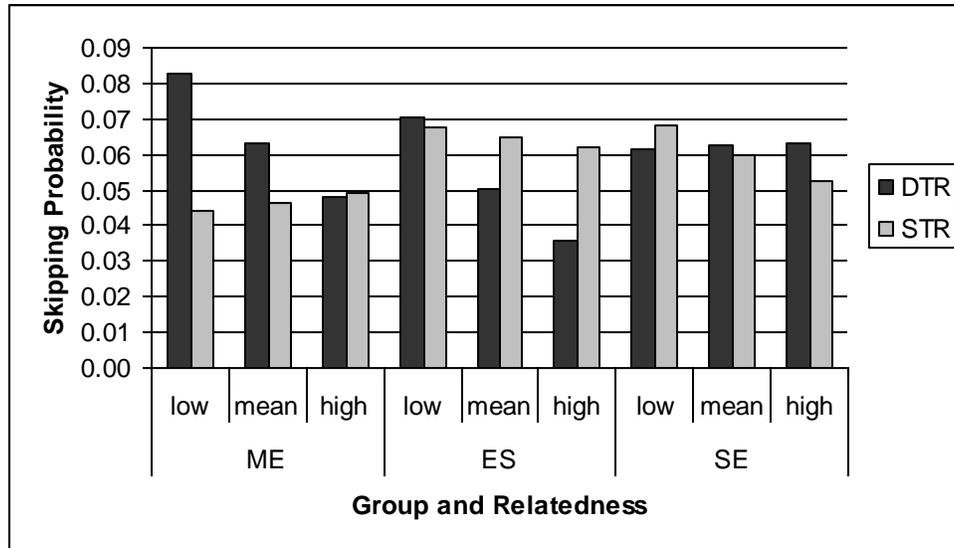
	SK	FFD	GD	GPT	TT	RO
Intercept	-2.697**	5.333**	5.579**	5.716**	6.028**	-2.154**
Participant Age	-0.008	0.000	0.000	0.001	0.001	0.015*
Mean Performance on IST sentences	7.422**	0.004**	0.003**	0.002**	0.002**	4.569**
Target Log KF frequency	-0.062	-0.004	-0.004	0.005	0.011	0.046
Target Length (in letters)	0.061	0.008±	-0.010	-0.020±	-0.029	-0.019
Target Part of Speech [verb]	-0.290	0.023	0.027	0.017	0.032	-0.002
Sentence Predictability Rating	0.419±	0.010	0.057±	0.047	-0.042	0.054
Sentence Un-naturalness Rating	-0.177*	0.004	0.007	0.011	0.053**	0.029
Log Form-Similarity (to the IST)	0.140	0.003	-0.023	0.005	-0.063±	-0.050
Semantic Relatedness (to the IST)	-0.192±	0.002	0.007	-0.007	-0.002	-0.082±
Group [ES]	-0.236	0.020	-0.019	-0.022	0.013	-0.014
Group [SE]	-0.013	0.017	0.035	0.018	-0.025	-0.294±

Condition [STR]	-0.320±	-0.013	-0.035±	0.003	-0.066**	0.197*
Condition [STR]:Group [ES]	0.585*	0.001	0.034	0.020	-0.001	-0.100
Condition [STR]:Group [SE]	0.279	0.007	0.006	-0.008	0.000	-0.067
Condition [STR]:Semantic Relatedness	0.229±	0.006	-0.007	-0.010	0.010	0.005
Group [ES]: Semantic Relatedness	-0.042	0.006	-0.025	-0.018	0.014	0.082
Group [SE]: Semantic Relatedness	0.201	-0.003	-0.005	0.003	-0.007	0.108
Condition [STR]: Group [ES]: Semantic Relatedness	-0.025	0.010	0.048*	0.048*	0.009	0.082
Condition [STR]: Group [SE]: Semantic Relatedness	-0.328±	-0.009	-0.003	0.001	0.019	-0.047

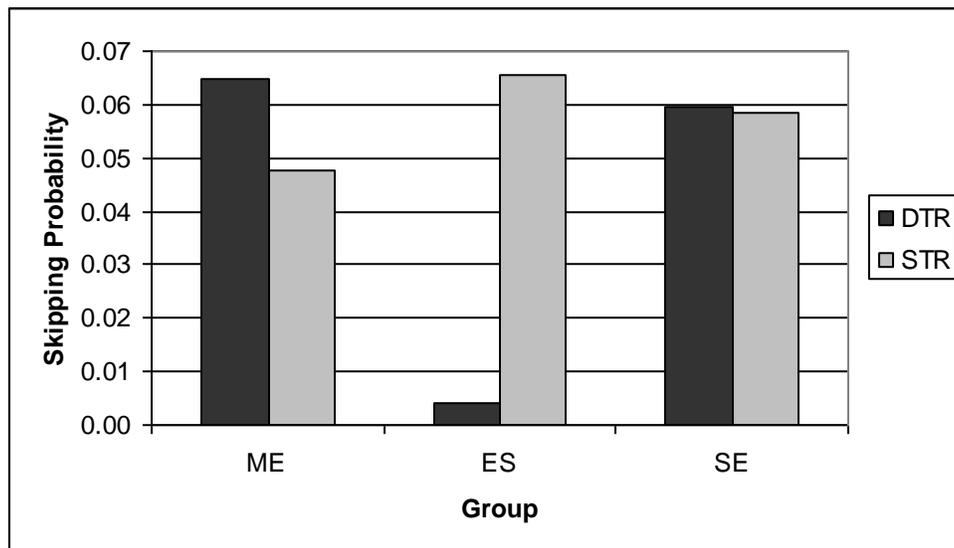
Note: ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

*Skipping:* Marginal effects of semantic relatedness and condition were qualified by a two-way interaction of relatedness and condition, and a three-way interaction of these factors with group. Specifically, as can be seen in Figure 19, ME speakers skipped the pre-target region more often in DTR sentences than in STR sentences when semantic relatedness was low. SE bilinguals, in contrast, skipped the pre-target region more often in DTR sentences when semantic relatedness was high. Additionally, the two-way interaction between condition and group suggest that the effect of condition was different for ES bilinguals compared to ME speakers, such that ES bilinguals skipped the pre-target region more often in STR sentences, whereas ME speakers skipped the pre-target region more often in DTR sentences (see Figure 19 and Figure 20). Note that the effects of condition on skipping of the pre-target region are surprising because sentences in DTR and STR conditions were identical up until the target region. There is some debate surrounding the possibility of  $n+2$  preview effects, such that information on the target word would be extracted from words preceding the pre-target region (e.g., Kliegl, Risse, & Laubrock,

2007; Rayner, Juhasz, & Brown, 2007) to allow these pre-target skips. If n+2 preview is not responsible for these differences, they could be due to chance.



**Figure 19.** Skipping probability of the pre-target region as a function of group, relatedness, and condition.



**Figure 20.** Skipping probability of the pre-target region as a function of group.

*First-Fixation Durations:* No relevant effects were significant.

*Gaze Durations:* GD on the pre-target region marginally varied by condition, with DTR sentences eliciting longer GDs than STR sentences, and this effect was qualified by a three-way

interaction among condition, group, and relatedness. As can be seen in Table 19, ME speakers exhibit longer durations on DTR sentences compared to STR sentences but ES bilinguals exhibit a different pattern, such that when relatedness is low, ES bilinguals exhibit longer durations on DTR sentences than on STR sentences, but when relatedness is high, ES bilinguals exhibit longer durations on STR than DTR sentences.

**Table 19.** Model estimates for mean performance in the pre-target region as a function of group, condition, and semantic relatedness.

Group	Relatedness Level	GD		GPT	
		DTR	STR	DTR	STR
ME	Low	261.97	255.93	307.04	312.57
	Mean	264.75	255.78	303.72	304.48
	High	267.57	255.62	300.43	296.59
ES	Low	267.19	250.72	308.94	298.05
	Mean	259.80	259.64	297.23	304.05
	High	252.60	268.88	285.96	310.18
SE	Low	273.36	269.79	311.08	313.80
	Mean	274.16	266.37	309.08	307.42
	High	274.96	263.00	307.09	301.16

These effects may be due to inflated durations on trials preceding a skip, because as described in the analyses of the target region, the same conditions (low relatedness DTR and high relatedness STR) also elicited more skips of the target region. To examine this possibility we analyzed the data after excluding trials on which the target region was skipped. Supporting this proposal, the effect of condition and its interaction with relatedness and group were no longer significant (see, Table 20).

**Table 20.** Coefficient estimates for the pre-target region excluding trials with target skips

	<b>SK</b>	<b>FFD</b>	<b>GD</b>	<b>GPT</b>	<b>TT</b>	<b>RO</b>
Intercept	-2.636**	5.335**	5.545**	5.686**	6.011**	-2.160**
Participant Age	-0.007	0.000	-0.001	0.001	0.001	0.017*
Mean Performance on IST sentences	9.472**	0.004**	0.003**	0.002**	0.002**	4.736**
Target Log KF frequency	-0.089	-0.005	-0.001	0.010	0.009	0.058
Target Length (in letters)	-0.007	0.007±	-0.004	-0.011	-0.027±	-0.007
Target Part of Speech [verb]	-0.270	0.024	0.027	0.025	0.036	0.048
Sentence Predictability Rating	0.354	0.015	0.086*	0.068±	-0.036	0.052
Sentence Un-naturalness Rating	-0.180*	0.000	0.001	0.011	0.054**	0.071
Log Form-Similarity (to the IST)	0.262	0.004	-0.018	-0.019	-0.060±	-0.213
Semantic Relatedness (to the IST)	-0.096	0.001	-0.003	-0.008	-0.001	-0.039
Group [ES]	-0.212	0.012	0.001	-0.007	0.017	-0.052
Group [SE]	-0.170	0.015	0.051	0.029	-0.019	-0.275
Condition [STR]	-0.302	-0.011	-0.024	0.001	-0.066**	0.193±
Condition [STR]:Group [ES]	0.499*	-0.012	0.023	-0.001	-0.012	-0.099
Condition [STR]:Group [SE]	0.281	-0.010	-0.007	-0.013	-0.015	0.011
Condition [STR]:Semantic Relatedness	0.134	0.007	0.009	0.004	0.015	-0.009
Group [ES]: Semantic Relatedness	-0.099	0.005	0.006	0.006	0.005	0.028
Group [SE]: Semantic Relatedness	-0.004	-0.002	-0.004	-0.005	-0.008	0.026
Condition [STR]: Group [ES]: Semantic Relatedness	0.058	0.003	0.017	0.011	-0.022	-0.011
Condition [STR]: Group [SE]: Semantic Relatedness	-0.202	-0.009	-0.004	-0.014	0.008	-0.145

Note: ±  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .001$

*Go-Past Time:* The three-way interaction among condition, group, and relatedness was reliable such that again ES bilinguals showed longer GPT on the pre-target region in DTR sentences when relatedness was low, but longer GPT in STR sentences when relatedness was high. As with the GD pattern, this interaction was not significant, however, after removal of trials with target skips (see Table 20), suggesting it was driven by inflated durations on trials preceding a skip of the target region.

*Total Time:* DTR sentences elicited significantly longer total viewing time on the pre-target region compared to STR sentences, but this effect did not vary as a function of group.

*Regressions Out:* The probability of regressing out of the pre-target region marginally increased as the semantic relatedness of the target word to the intended IST decreased. SE bilinguals tended to regress out of the pre-target region significantly less often than ME speakers. And finally, regressions out of the pre-target region were significantly more likely in STR sentences compared to DTR sentences, but this effect was not qualified by a group or relatedness interaction.

## APPENDIX F

### RELATENDNESS RATINGS OF EXPRESSION PAIRS FOR EXPERIMENT 2

#### F.1 METHOD

##### F.1.1 Participants

Twenty monolingual English speakers (mean age 19.55 years,  $SD=2.11$ ; 9 females) completed this rating task toward credit in an Introductory Psychology class. They were not exposed to languages other than English before age 10, and all but one indicated being exposed to English at least 80% of the time at the time of testing. Three participants indicated some exposure to Spanish, but rated their proficiency and use to be less than 4.5 on a 1-10 scale, with 1 indicating ‘not proficient’ and ‘hardly ever use’, and 10 indicates ‘very proficient’ and ‘always use’. Data from 4 participants were replaced because they were exposed to languages other than English before age 10 or rated their proficiency in Spanish to be above 4.5.

### **F.1.2 Materials**

A set of 185 ambiguous English words were selected from available within-language ambiguity research (e.g., Klepousniotou et al., 2008; Nelson et al., 1980; Twilley et al., 1994), such that some of them have a single translation in Spanish that captures both meanings/senses of the English word ('joint translation' condition, e.g., 'cuerpo' captures both the administrative and the biological meanings of the word 'body'; 84 items), and some have two Spanish translations ('split translation' condition, each encompassing one of the senses/meanings of the English word; 101 items). Translations in Spanish were determined by consulting two highly proficient Spanish-English bilinguals.

Two modifiers were selected for each ambiguous word, each highlighting a different sense or meaning of the word. For example, 'human body–administrative body'. Two versions were constructed such that the order of the modified expressions was altered from one version to the next.

### **F.1.3 Procedure**

Participants completed this rating task via a web-based interface. They were presented with a randomized list of the expression pairs (e.g., 'human body–administrative body'), and were asked to rate how similar in meaning the two senses of the ambiguous word were based on the modifiers with which they were presented, on a scale from 1 ('completely different') to 7 ('exactly the same'). Participants were encouraged to use the full range of the scale and to rate the expressions in the order in which they were presented. Participants were provided with one example, and were informed that this was not a test of their knowledge but rather that we were

interested in their intuitions. Following the rating task, participants completed a detailed language history questionnaire (see Table 14).

Based on these norms a meaning similarity score was computed for each ambiguous item based on the average similarity rating in the two versions. These meaning similarity scores served as predictors in the analyses of the semantic judgment task (see Chapter 3).

## APPENDIX G

### CRITICAL EXPRESSION PAIRS EXPERIMENT 2

**Table 21.** Critical expression pairs Experiment 2. Rel: Mean Semantic Relatedness Ratings (1-7 scale)

Condition	Expression Pair	Rel
joint	rock band -- rubber band	1.05
joint	organ donor -- pipe organ	1.10
joint	infinitely patient -- cancer patient	1.25
joint	savings bank -- riverbank	1.30
joint	degrees Fahrenheit -- college degrees	1.40
joint	investment capital -- capital letter	1.45
joint	house plant -- power plant	1.50
joint	right angle -- reporter's angle	1.60
joint	bird's wing -- east wing	1.60
joint	rhythmic movement -- civil-rights movement	1.60
joint	movie admission -- false admission	1.65
joint	mental state -- southern state	1.65
joint	dinner reservation -- Indian reservation	1.70
joint	master key -- key problem	1.75
joint	kitchen cabinet -- presidential cabinet	1.75
joint	American Revolution -- axial revolution	1.79

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
joint	dry skin -- dry wine	1.95
joint	earned interest -- extracurricular interest	2.00
joint	secret passage -- literary passage	2.05
joint	introductory course -- collision course	2.08
joint	pharmaceutical company -- pleasant company	2.10
joint	bread mold -- plaster mold	2.15
joint	first impression -- shoe impression	2.30
joint	function properly -- social function	2.33
joint	figure caption -- body figure	2.41
joint	left foot -- foot long	2.45
joint	first act -- act carefully	2.45
joint	theoretical model -- fashion model	2.45
joint	television volume -- volume measure	2.45
joint	large head -- organization's head	2.50
joint	flea market -- housing market	2.55
joint	informal atmosphere -- polluted atmosphere	2.55
joint	human body -- administrative body	2.65
joint	whole grain -- coarse grain	2.65
joint	shining star -- movie star	2.75
joint	waffle cone -- traffic cone	2.75
joint	clotted blood -- royal blood	2.90
joint	gossip column -- support column	2.90
joint	right hand -- helping hand	3.05
joint	large object -- direct object	3.07
joint	true nature -- mother nature	3.10
joint	TV guide -- tour guide	3.12

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
joint	military resistance -- electric resistance	3.15
joint	impulsive reaction -- chemical reaction	3.20
joint	military operation -- mathematical operation	3.25
joint	good condition -- first condition	3.35
joint	compact disc -- compact car	3.35
joint	police barrier -- language barrier	3.39
joint	inherited fortune -- good fortune	3.40
joint	reason why -- use reason	3.50
joint	olive oil -- motor oil	3.70
joint	multiple-choice examination -- thorough examination	3.80
joint	embarrassing position -- fetal position	3.85
joint	juicy orange -- bright orange	3.95
joint	essay title -- formal title	4.00
joint	open space -- outer space	4.00
joint	storm cloud -- mysterious cloud	4.05
joint	charitable contribution -- scholarly contribution	4.05
joint	feature film -- 35mm film	4.06
joint	express concern -- national concern	4.10
joint	pancake breakfast -- lonely breakfast	4.20
joint	artificial intelligence -- military intelligence	4.20
joint	cough medicine -- veterinary medicine	4.29
joint	hot lunch -- ladies lunch	4.40
joint	dollar bill -- weak dollar	4.45
joint	beautiful design -- architectural design	4.50
joint	bad dream -- childhood dream	4.50
joint	beating heart -- broken heart	4.52

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
joint	big class -- boring class	4.54
joint	state park -- amusement park	4.55
joint	security guard -- coast guard	4.71
joint	main road -- icy road	4.80
joint	teen magazine -- glossy magazine	4.90
joint	marinated lamb -- baby lamb	4.95
joint	history article -- popular article	5.00
joint	best-selling book -- leather-bound book	5.05
joint	phone message -- urgent message	5.10
joint	light dinner -- formal dinner	5.10
joint	planet earth -- fertile earth	5.10
joint	brute force -- military force	5.15
joint	fried chicken -- clucking chicken	5.23
joint	daily newspaper -- shredded newspaper	5.25
joint	cotton dress -- cotton thread	5.30
joint	major difference -- difference between	5.33
split	straight line -- checkout line	3.55
split	lean meat -- lean towards	1.10
split	red rose -- rose above	1.15
split	diamond ring -- loud ring	1.15
split	express train -- train dogs	1.20
split	private plane -- flat plane	1.20
split	soccer ball -- formal ball	1.20
split	head count -- Count Dracula	1.20
split	plaster cast -- news cast	1.20
split	bed rest -- all the rest	1.25

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
split	breached contract -- contract the flu	1.30
split	properly spell -- magical spell	1.30
split	excessively mean -- mathematical mean	1.30
split	grizzly bear -- bear down	1.31
split	spring rain -- coiled spring	1.35
split	white lie -- lie down	1.35
split	bright light -- light weight	1.35
split	good pick -- ice pick	1.35
split	ironing board -- school board	1.35
split	night club -- wooden club	1.40
split	suitcase -- criminal case	1.40
split	iron clothes -- iron supplement	1.40
split	pop rock -- volcanic rock	1.40
split	hair comb -- rooster's comb	1.45
split	cookie jar -- jar suddenly	1.45
split	wrist watch -- watch television	1.45
split	Sunday drive -- internal drive	1.50
split	wood bat -- vampire bat	1.50
split	match stick -- stick around	1.50
split	half pound -- pound loudly	1.55
split	cardboard box -- kick box	1.55
split	left arm -- fire arm	1.60
split	large square -- square root	1.60
split	stand up -- night stand	1.65
split	wooden block -- mental block	1.65
split	love letter -- letter grade	1.70

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
split	birthday party -- democratic party	1.70
split	solid ground -- finely ground	1.75
split	barely try -- try in court	1.79
split	mini skirt -- skirt around	1.80
split	contact lens -- human contact	1.80
split	tsunami wave -- wave good-bye	1.80
split	power drill -- practice drill	1.85
split	stop sign -- sign here	1.85
split	hang up -- hang out	1.90
split	hall pass -- barely pass	1.90
split	sex appeal -- legal appeal	1.90
split	laser beam -- wood beam	1.90
split	oscillating fan -- biggest fan	1.95
split	red lip -- protective lip	2.00
split	fly a plane -- house fly	2.00
split	freezing cold -- common cold	2.00
split	undeniable proof -- proofread	2.05
split	double-sided tape -- video tape	2.05
split	don't like -- exactly like	2.10
split	mastermind -- mind your manners	2.10
split	trust me -- trust fund	2.10
split	under cover -- bed cover	2.10
split	best-selling novel -- completely novel	2.10
split	sharp point -- point a finger	2.15
split	shoot a turkey -- photo shoot	2.20
split	camp fire -- gun fire	2.21

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
split	suspension bridge -- playing bridge	2.25
split	true story -- second story	2.25
split	dash mark -- high mark	2.25
split	winter coat -- clear coat	2.30
split	indoor tracks -- deer tracks	2.30
split	economic power -- electric power	2.35
split	jury trial -- clinical trial	2.35
split	quick finish -- glossy finish	2.40
split	grocery store -- store grain	2.46
split	child's play -- play music	2.50
split	cruise ship -- ship across	2.50
split	honorable cause -- might cause	2.60
split	dinner date -- expiration date	2.60
split	kick the ball -- kick the habit	2.65
split	charitable foundation -- strong foundation	2.70
split	fall down -- fall leaves	2.75
split	car race -- human race	2.90
split	control panel -- advisory panel	2.90
split	summer home -- funeral home	2.95
split	don't approach -- new approach	3.00
split	bow tie -- tie tightly	3.15
split	touching scene -- panoramic scene	3.15
split	winter term -- long- term	3.20
split	phone call -- call out	3.25
split	combination lock -- please lock	3.25
split	stained glass -- empty glass	3.35

<b>Condition</b>	<b>Expression Pair</b>	<b>Rel</b>
split	long dress -- dress appropriately	3.50
split	fair chance -- chance of a lifetime	3.55
split	primary aim -- aim higher	3.60
split	guest room -- adequate room	3.70
split	pay rent -- rent a movie	3.80
split	drink quickly -- cold drink	3.80
split	laundry pile -- pile up	3.95
split	slow answer -- correct answer	4.00
split	sensitive issue -- previous issue	4.04
split	in the air -- fresh air	4.10
split	burn quickly -- third-degree burn	4.15
split	foreign country -- wine country	4.15
split	profitable business -- legitimate business	5.70

## **APPENDIX H**

### **TASK INSTRUCTIONS**

#### **H.1 SENTENCE READING TASK**

Instructions were taken from previous research on semantic anomaly (e.g., Patson & Warren, 2010).

“In this experiment, you will be shown a series of sentences, one at a time. After some sentences, you will be asked a "yes" or "no" comprehension question about the sentence that you have just read. Please read the sentences as naturally as possible.

Please let the experimenter know that you are ready to begin the experiment. If you have any questions about what you are supposed to do, ask the experimenter at this time.”

#### **H.2 SENTENCE NATURALNESS RATING TASK**

Instructions were taken from previous research on semantic anomaly (e.g., Patson & Warren, 2010).

“In the following questionnaire, sentences will vary in how natural they sound.

Your task is to rate how natural each sentence sounds on a scale of 1 (very natural) to 7 (very unnatural).

Of course, there will also be sentences that fall between the extremes. Please use the full range of the scale. If you are not sure how to rate a particular sentence, it is appropriate to guess or follow your first instinct. Keep in mind that this is not a test of your knowledge; we are simply interested in your intuition.

Thank you for your participation.

Examples:

1. Joe climbed the tree to get down the ball that was stuck....

This sentence is very natural (it would probably be rated a 1 or a 2).

Explanation: It is pretty normal to climb a tree to get a ball that is stuck in it, so this sentence sounds natural.

2. Joe climbed the Cathedral of Learning to get down the ball that was stuck....

This sentence is very unnatural (it would probably be rated a 5 or a 6).

Explanation: The Cathedral of Learning is very big, so it would be very unlikely that someone would climb it to get a ball stuck on top of it.

3. Joe climbed the tree to get down the dog that was stuck...

This sentence is somewhere in the middle of the ratings (it would probably be rated a 3 or a 4).

Explanation: Although dogs don't normally get stuck in trees, you can easily imagine someone climbing a tree to rescue a dog that can't get down.

Please rate the sentences in the order in which they appear in the list. Please do not change your responses or go back to a previous sentence.”

### H.3 SEMANTIC RELATEDNESS TASK

“Ambiguous words have more than one meaning/sense. For example, the word 'bark' refers to the sound a dog makes and to the outer layer of a tree. In this task, you will be presented with ambiguous words along with modifiers of their two senses, at the same time

For instance:

fall down -- fall leaves

Your task is to decide as quickly and accurately as possible if the two senses of the ambiguous word are related in meaning.

Press the 'Yes' key if you think they are related in meaning. Press the 'No' key if you think they are not related.

For instance:

tree bark -- dog bark

are not related in meaning, but:

personal check -- cashier's check

are related in meaning, because both expressions refer to the monetary sense of the word check.

In each trial, a fixation cross (+) will be presented at the center of the screen, followed by a pair of expressions.

Remember, press 'yes' if you think the meanings are related, and 'no' if you think they are not. As soon as you respond, the expressions will disappear from the screen and be replaced by the cross again.

If you are not sure, it is ok to guess or follow your first instinct. If you make a mistake, simply continue to the next trial. If you have any questions, ask the experimenter now.”

#### **H.4 LEXICAL DECISION**

“In this experiment, two letter strings will appear one after the other in each trial, at the center of the screen. The first letter string will always be a real English [Spanish] word.

Your task is to read both letter strings and to decide if the second letter string is a real English [Spanish] word. Please respond as quickly and accurately as possible as soon as you see the second letter string. press the Yes key if the second letter string IS a real word in English [Spanish]. press the No key if the second letter string IS NOT a real word in English [Spanish]. Please keep your right index finger on the Yes key and your left index finger on the No key at all times. Please read both letter strings, but respond only to the second string. If you are not sure, it is ok to guess. As soon as you have responded, the letter string will disappear from the screen, and the next trial will begin with a fixation cross. You will have some practice trials to become comfortable with this task. You will be notified when the practice trials are over. If you have any questions, please ask the experimenter now.”

#### **H.5 PICTURE NAMING**

“At the fixation cross (+) press the middle key of the response pad to see the picture.

When you see the picture, say the name of the picture in English [Spanish] out loud.

Remember to speak loudly and clearly. If you do not know the name of the picture, say "no" out loud. If you are uncertain of the name of the picture, it's okay to guess. Please respond as quickly and accurately as you can. If you have any questions, please ask the experimenter now.”

## H.6 OPERATION-WORD SPAN

“In this task, you will see a series of simple arithmetic operations, some of which have correct answers and some of which do not. For example:

$(4 / 2) + 6 = 8$  is correct, but,

$(3 / 1) - 2 = 5$  is incorrect.

After each operation, your task is to decide if the answer given was correct or not. If you think it was correct, press 'YES' with your right finger when the question mark appears. If you think the answer was not correct, press 'NO' with your left finger when the question mark appears. Please try to respond as quickly and accurately as possible, after the question mark.

In addition, you will periodically be asked to recall and type in some words that are presented after the operations (all of the words will be in English [Spanish]).

The operations will occur in sets ranging in size from 2 to 6, with three sets of each size. After each set of operations, the word "RECALL" will flash on the screen. At that point, you should type as many of the words as you can remember from that set. Please try to type the words in the order in which they appeared. When you have finished typing, press ESC to begin the next set.

There will be two practice sets, after which you will be notified that the experiment is about to begin.

Remember, you should respond as quickly as possible because there is a time limit. If you do not answer while the question mark is on the screen, your response will automatically be considered incorrect. Please make your response ONLY when the question mark is on the screen. Answer the operations as accurately as possible.”

## H.7 RAVEN'S MATRICES

“Consider the set of patterns within the box at the top of the page. Each row and each column become increasingly complex as you move from left to right and from top to bottom, respectively. To complete the group (and fill in the blank) you must choose (by clicking on) the next pattern in the series, the one that satisfies the series of both the bottom row and the rightmost column simultaneously. Choose your answer at the bottom of the page. The first question shown will be an example.”

## APPENDIX I

### LANGUAGE HISTORY QUESTIONNAIRE

This questionnaire is designed to give us a better understanding of your language experience, as well as your general reading and computer experience.

We ask that you be as accurate and thorough as possible when answering the following questions and thank you for your participation in this study.

**Sex:** \_\_\_\_\_ **Age (in years)** \_\_\_\_\_ **Handedness :** \_\_\_\_\_ **Native country** \_\_\_\_\_

**Do you have normal or corrected-to-normal vision?** \_\_\_\_\_ **If not, please describe** \_\_\_\_\_

1.) Please list all the languages you know **in order of dominance**:

1 \_\_\_\_\_ 2 \_\_\_\_\_ 3 \_\_\_\_\_ 4 \_\_\_\_\_ 5 \_\_\_\_\_

2.) Please list all the languages you know **in order of acquisition** (your native language first). Also specify the **age** in years at which you **began** to learn the language and the **context** in which you learned it. For example, "English, birth, home". Include all languages to which you have been exposed, although you may never have had formal training in them and may not be able to read, speak or write them.

<b>Language</b>	1 _____	2 _____	3 _____	4 _____	5 _____
<b>Age (when began)</b>					
<b>Learning Situation</b>					

3.) Please list what percentage of the time your are **currently** and **on average** exposed to each language (Your percentage should add up to 100%)

<b>Language</b>					
<b>Percentage %</b>					

4.) When choosing to read a text available in all your languages, in what percentage of cases would you choose to **read** it in each of your languages? Assume that the original was written in another language, which is unknown to you. (Your percentage should add up to 100%)

<b>Language</b>					
<b>Percentage %</b>					

5.) When choosing to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to **speak** each language? Please report percent of total time. (Your percentage should add up to 100%)

<b>Language</b>					
<b>Percentage %</b>					

**NOTE: IF YOU KNOW ENGLISH & SPANISH ANSWER QUESTIONS 6-11 ABOUT THOSE LANGUAGES, even if another language is your first/second language.**

6.) Please rate your **first language** \_\_\_\_\_ proficiency on a ten-point scale. (1= not proficient, 10= very proficient)

	Not proficient									Very proficient
Reading	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Conversation Fluency	1	2	3	4	5	6	7	8	9	10
Oral Comprehension	1	2	3	4	5	6	7	8	9	10

7.) Please use the following scale to describe your current use of your **first language** \_\_\_:

	Hardly ever									Always
Speaking	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Reading	1	2	3	4	5	6	7	8	9	10
Listening to radio/music	1	2	3	4	5	6	7	8	9	10
Watching TV	1	2	3	4	5	6	7	8	9	10
Other (describe)___	1	2	3	4	5	6	7	8	9	10

8.) With respect to your **second language** \_\_\_\_\_ , please indicate

**Number of years** you have studied the language: \_\_\_\_\_

**setting(s)** in which you have had experience with the language (i.e., classroom, with friends, foreign country...) \_\_\_\_\_

9.) Please rate your **second language** \_\_\_\_\_ proficiency on a ten-point scale.(1= not proficient, 10= very proficient)

	Not proficient									Very proficient
Reading	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Conversation Fluency	1	2	3	4	5	6	7	8	9	10
Oral Comprehension	1	2	3	4	5	6	7	8	9	10

10.) Please use the following scale to describe your current use of your **second language** \_\_\_\_\_:

	Hardly ever									Always
Speaking	1	2	3	4	5	6	7	8	9	10
Writing	1	2	3	4	5	6	7	8	9	10
Reading	1	2	3	4	5	6	7	8	9	10
Listening to radio/music	1	2	3	4	5	6	7	8	9	10
Watching TV	1	2	3	4	5	6	7	8	9	10
Other (describe)_____	1	2	3	4	5	6	7	8	9	10

11.) Have you ever lived in a place where the dominant language spoken was your second language (i.e. immersion)? Yes/no

**If you have**, how long were you immersed in your second language environment? \_\_\_\_\_

How much time has passed since your most recent immersion experience? \_\_\_\_\_

12.) How many years of formal education do you have? \_\_\_\_\_ Please check your highest education level (Less than high school/High school/Professional training/Some college/college/some graduate school/masters/PhD,MD,JD)

13.) How many years of formal education did your **mother** have before you turned 18? \_\_\_\_\_ Please check her highest education level (Less than high school/High school/Professional training/Some college/college/some graduate school/masters/PhD,MD,JD/ Unknown)

14.) How many years of formal education did your **father** have before you turned 18? \_\_\_\_\_ Please check her highest education level (Less than high school/High school/Professional training/Some college/college/some graduate school/masters/PhD,MD,JD/ Unknown)

15.) Please list the language(s) the following people speak.

Mother: \_\_\_\_\_

Father: \_\_\_\_\_

Spouse / Closest friend: \_\_\_\_\_

16.) Please use the following scale to describe your current attitude toward **reading**? 1-10

(1=very negative; 10=very positive) \_\_\_\_\_

17.) Please use the following scale to describe how much **reading** you do in the following contexts:

	None									A great deal
Pleasure	1	2	3	4	5	6	7	8	9	10
Work	1	2	3	4	5	6	7	8	9	10
School	1	2	3	4	5	6	7	8	9	10
Other (describe)___	1	2	3	4	5	6	7	8	9	10

18.) Have you ever been diagnosed with a **reading disability**? Yes/no **Describe** \_\_\_\_

19.) Please rate your **general computer skill** on a ten-point scale.(1= not proficient, 10= very proficient) \_\_\_\_\_

20.) Please use the following scale to describe your current **use of computers** in the following contexts:

	Hardly ever									Always
Home	1	2	3	4	5	6	7	8	9	10
Work	1	2	3	4	5	6	7	8	9	10
School	1	2	3	4	5	6	7	8	9	10
Other (describe)___	1	2	3	4	5	6	7	8	9	10

21.) Please rate the extent to which you play **fast-action video games** on a ten-point scale. (1=hardly ever, 10=every day)

22.) Please use the following scale to describe your **skill in playing fast-action video games**

23.) Is there anything else about your language/general background that you would like to comment on? Please feel free to make comments about things which were not covered on this questionnaire. \_\_\_\_\_

24.) What do you think we were studying in this experiment? \_\_\_\_\_

## APPENDIX J

### VOCABULARY POST-TEST QUESTIONNAIRE

#### Vocabulary Checklist

Please read through the following list of Spanish-English translation pairs. Please **mark an X** next to any translation pairs you are not familiar with. Note that some Spanish words correspond to 2 English translations, and vice versa, but judge each pair separately. If you are familiar with one translation-pair but not the other, mark only the one you are not familiar with. You should mark an X if you do not know the English or the Spanish word. Thank you for your participation.

escritura – deed		destino - destiny		globo - globe
premio – award		papa - potato		silla - chair
pelea – quarrel		vela - candle		excitación - excitement
plato – dish		hora - hour		balance - balance
ejercicio – drill		cámara - chamber		apuntar - point
cresta - cockscomb		razón - cause		cura - cure
moneda – money		suerte - chance		bosque - forest
techo – ceiling		cabo - end		manejar - drive
hora – time		cólera - anger		descubrimiento - discovery
plato – plate		suerte - luck		borde - edge

mundo – world		pata - leg		pico - peak
equilibrio - equilibrium		vacación - vacation		anuncio - sign
razón – reason		discusión - argument		investigación - investigation
sierra – saw		cargar - charge		marca - mark
vestir – dress		bote - ship		excitación - agitation
copa – goblet		dirección - direction		capacidad - competence
juicio – judgment		calor - warmth		ayudar - assist
saltar – leap		cola - tail		otoño - fall
culpa – fault		cólera - cholera		manejar - manage
tomar – drink		serpiente - serpent		vela - sail
conducir – drive		ejército - army		cocinero - cook
respuesta - answer		lástima - shame		planta - plant
negar – negate		importar - mind		borde - border
moneda – coin		película - film		lengua - tongue
pista – path		corte - court		oferta - offer
carne – flesh		tienda - store		cola - glue
trasladar - transfer		sarcasmo - sarcasm		banco - bench
lastimar – harm		noticia - news		decepcionar - disappoint
intento - intention		vehículo - vehicle		carne - meat
serpiente – snake		tomar - take		bala - bullet
mujer – woman		silla - seat		cocinero - chef
saltar – jump		apuntar - note		prueba - proof
cuento – story		vestir - wear		sofá - couch
fuerza – strength		respuesta - response		prueba - test
noticia – notice		armario - dresser		muñeca - wrist
confianza - confidence		hablar - talk		denegar - reject
carta – card		invertir - invert		cabo - cape

intento – attempt		dirección - address		pieza - room
pila – pile		invento - invention		mujer - wife
alivio – relief		banco - bank		lástima - pity
decepcionar - deceive		piedra - stone		casa - house
camisa – shirt		carta - letter		fuerza - force
peligro – danger		marca - brand		dibujar - draw
lastimar – hurt		película - movie		llama - flame
calle – road		maldición - swear		pieza - piece
casa – home		armario - closet		negar - deny
tocar – touch		muñeca - doll		sierra - range
obligación – duty		necesidad - necessity		otoño - autumn
conejo – bunny		comida - meal		reloj - clock
cura – minister		don - sir		asistencia - attendance
charlar – chat		carrera - career		piedra - rock
alivio - alleviation		escritura - writing		dibujar - sketch
tienda – tent		globo - balloon		confianza - trust
boleto – ticket		desear - want		asistencia - assistance
investigación - research		equilibrio - balance		anuncio - announcement
brindis – toast		pata - duck		ayudar - help
capacidad - ability		denegar - refuse		juicio - trial
bala – bale		atracar - dock		empresa - business
importar – care		planta - floor		corte - cut
cámara – camera		cargar - carry		lengua - language
invento - fabrication		comida - food		maldición - curse
necesidad – need		trasladar - move		descubrimiento - finding
premio – prize		carrera - race		desear - wish
sarcasmo – irony		hablar - speak		punto - point

	cita - appointment		discusión - discussion		vehículo - automobile
	mundo – earth		culpa - guilt		reloj - watch
	pista – clue		cuento - tale		juntar - gather
	tocar – play		escoger - pick		brindis - cheer
	sofá – sofa		charlar - talk		pico - beak
	pelea – fight		cresta - crest		llama - llama
	don – talent		oferta - sale		techo - roof
	balance – scale		calle - street		ejército - military
	escoger – choose		punto - period		obligación - obligation
	empresa - company		vacación - holiday		peligro - trouble
	conejo – rabbit		pila - battery		camisa - blouse
	papa – pope		calor - heat		bosque - wood
	ejercicio - exercise		título - title		juntar - join
	cita – citation		invertir - invest		atracar - rob
	bote – boat		título - degree		boleto - voucher
	copa – treetop		conducir - conduct		destino - fate

## BIBLIOGRAPHY

- Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2008). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, *60*, 270-290.
- Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language*, *53*, 60-80.
- Arêas Da Luz Fontes, A. B., & Schwartz, A. I. (2010). On a different plane: Cross-language effects on the conceptual representations of within-language homonyms. *Language and Cognitive Processes*, *25*, 508-532.
- Armstrong, B. C., & Plaut, D. C. (2008, July). *Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence*. Paper presented at the 30th Annual Conference of the Cognitive Science Society, Hillsdale, NJ.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. New York: Cambridge University Press.
- 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*, 390-412.
- Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 827-845.
- Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*, 235-260.
- Berreta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research*, *24*, 57-65.
- Boada, R., Sánchez-Casas, R., García-Albea, J. E., Gaviln, J. M., & Ferr, P. (2009, July). *Effect of number of translations and cognate status in translation recognition performance of proficient bilinguals*. Paper presented at the International Symposium on Bilingualism. Utrecht, The Netherlands.

- Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition*, 30, 225-251.
- Brysbaert, M., Drieghe, D. & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed), *Cognitive processes in eye guidance* (pp 53-78). Oxford, UK: Oxford University Press.
- Chwilla, D. J., & Kolk, H. H. J. (2003). Event-related potential and reaction time evidence for inhibition between alternative meanings of ambiguous words. *Brain and Language*, 86, 167-192.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, 12, 769-786.
- Cook, V. (2003). *Effects of the second language on the first*. Clevedon, U. K.: Multilingual Matters.
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50, 491-511.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665-671.
- de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1001-1018.
- Degani, T., & Tokowicz, N. (2010a). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition*, 13, 299-314.
- Degani, T., & Tokowicz, N. (2010b). Semantic ambiguity within and across languages: An integrative review. *The Quarterly Journal of Experimental Psychology*, 63, 1266-1303.
- Degani, T., Eddington, C. M., Tokowicz, N., & Prior, A. (2009, November). *Comparing within-language and across-language word ambiguity*. Poster presented at the Fiftieth Annual Meeting of the Psychonomic Society, Boston, MA.
- Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, 23, 18-28.
- DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J.F. Kroll & A.M.B. De Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp.88-108).Oxford: Oxford University Press.

- Dijkstra, T., & 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): 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, 175-197.
- Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, 8, 221-238.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429-446.
- Durkin, K., & Manning, J. (1989). Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses. *Journal of Psycholinguistic Research*, 18, 577-612.
- Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 10, 101-116.
- Eddington, C. M., & Tokowicz, N. (2011). *Examining English-German translation ambiguity using primed translation recognition*. Manuscript in preparation.
- Eddington, C. M., Degani, T., & Tokowicz, N. (2011). *English and German translation norms: Examining ambiguity and semantic variability of translations across proficiency levels*. Manuscript in revision.
- Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, 28, 105-144.
- Edwards, J. V. (2004). Foundations of bilingualism. In T. K. Bhatia & W. C. Ritchie (Eds.), *The Handbook of Bilingualism* (pp. 7-31). Malden, MA: Blackwell.
- Elston-Güttler, K. E. (2000). *An enquiry into cross-language differences in lexical-conceptual relationships and their effects on L2 lexical processing*. Unpublished doctoral dissertation, University of Cambridge, England.
- Elston-Güttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, 52, 256-283.
- Elston-Güttler, K. E., & Williams, J. N. (2008). First language polysemy affects second language meaning interpretation: Evidence for activation of first language concepts during second language reading. *Second Language Research*, 24, 167-187.
- Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005a). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research*, 25, 57-70.

- Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005b). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*, 1593-1610.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*, 181-200.
- Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology A*, *50*, 119-148.
- Gollan, T. H., Montoya, R. I., & Fennema-Notestine, C. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition*, *33*, 1220-1234.
- Gorfein, D. S. (2001). *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity*: American Psychological Association.
- Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*, 265-274.
- Hohenstein, J., Eisenberg, A. N. N., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition*, *9*, 249-261.
- Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics*, *32*, 223-262.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*, 434-446.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47-77.
- Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*, 617-637.
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *Modern Language Journal*, *88*, 416-432.
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye movement investigation. *Language Learning*, *59*, 503-535.
- Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 601-609.
- Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*, 259-282.

- Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language*, 47(4), 548-570.
- Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20, 1-24.
- Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1534-1543.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262-284.
- Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n+2. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1250-1255.
- 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, 149-174.
- Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, 23(6), 701-717.
- Laxén, J., & Lavour, J. (2010). The role of semantics in translation recognition: effects of number of translations, dominance of translations and semantic relatedness of multiple translations. *Bilingualism: Language and Cognition*, 13, 157-183.
- Lenneberg, E. H. (1967). *Biological foundations of language*. Oxford: Wiley.
- 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, 381.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 49-67). Oxford: Oxford University Press.
- Marian, V., & Kaushanskaya, M. (2007). Cross-linguistic transfer and borrowing in bilinguals. *Applied Psycholinguistics*, 28, 369-390.
- 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, 940-967.

- Masson, M. E., & Freedman, L. (1990). Fluent identification of repeated words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 355-373.
- Michael, E. B., Tokowicz, N., Degani, T. & Smith, C. J. (2011). Individual differences in the ability to resolve translation ambiguity across languages. *Vigo International Journal of Applied Linguistics*, *8*, 79-97.
- Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., & Kroll, J. F. (2011). When deaf signers read English: Do written words activate their sign translations? *Cognition*, *118*, 286-292.
- Murray, W. S. (2006). The nature and time course of pragmatic plausibility effects. *Journal of psycholinguistic research*, *35*, 79-99.
- Nelson, D. L., McEvoy, C. L., Walling, J. R., & Wheeler, J. W. (1980). The University of South Florida homograph norms. *Behavior Research Methods & Instrumentation*, *12*, 16-37.
- Nelson, J. (2010). Reading skill and components of word knowledge affect eye movements during reading. Unpublished doctoral dissertation, University of Pittsburgh, PA.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, *9*, 225-236.
- Patson, N. D., & Warren, T. (2010). Eye movements when reading implausible sentences: Investigating potential structural influences on semantic integration. *The Quarterly Journal of Experimental Psychology*, *63*, 1516-1532.
- Pavlenko, A., & Jarvis, S. (2002). Bidirectional transfer. *Applied Linguistics*, *23*, 190-214.
- Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*, 1029-1038.
- Prior, A., Wintner, S., MacWhinney, B., & Lavie, A. (2010). Translation ambiguity in and out of context. *Applied Psycholinguistics*, *32*, 93-111.
- Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, *18*, 97-109.
- R development core team (2007). *R: A language and environment for statistical computing*. Vienna: R foundation for Statistical Computing, <http://www.R-project.org>
- Raven, J. C. (1960). *Guide to the Standard Progressive Matrices*. London: H K Lewis.
- Rayner, K. & McConkie, G.W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829-837.

- Rayner, K., Juhasz, B. J., & Brown, S. J. (2007). Do readers obtain preview benefit from word n+2? A test of serial attention shift versus distributed lexical processing models of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 230-245.
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 514-528.
- Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1290-1301.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125-157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445-526.
- 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.
- Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 243-256.
- Schwartz, A. I., & Arêas Da Luz Fontes, A. B. (2008). Cross-language mediated priming: Effects of context and lexical relationship. *Bilingualism: Language and Cognition*, *11*, 95-110.
- Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (2000). LEXESP: Léxico informatizado del español. Barcelona: Edicions de la Universitat de Barcelona.
- Simpson, G. B. (1981). Meaning Dominance and Semantic Context in the Processing of Lexical Ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*, 120-136.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*, 174-215.
- Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530-12535.
- Tokowicz, N. (1997). *Reevaluating concreteness effects in bilingual translation*. Unpublished Master's Thesis, The Pennsylvania State University, PA.

- Tokowicz, N., & Degani, T. (2010). Translation ambiguity: Consequences for learning and processing. In B. VanPatten & J. Jegerski (Eds.), *Research on second language processing and parsing*. Amsterdam: John Benjamins.
- Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, 22, 727-779.
- Tokowicz, N., Kroll, J. F., de Groot, A. M. B., & van Hell, J. G. (2002). Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments and Computers*, 34, 435-451.
- Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, 7, 255-272.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- Twilley, L. C., Dixon, P., Taylor, D., & Clark, K. (1994). University of Alberta norms of relative meaning frequency for 566 homographs. *Memory and Cognition*, 22, 111-126.
- van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, 64, 88-107.
- van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1, 193-211.
- van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review*, 9, 780-789.
- Warren, T. (in press). The influence of plausibility and anomaly on eye movements in reading. In I. Gilchrist, S. Everling, & S. Liversedge (Eds.), *The Oxford Handbook on Eye Movements*.
- Warren, T., & McConnell, K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin and Review*, 14, 770-775.
- Warren, T., McConnell, K., & Rayner, K. (2008). Effects of context on eye movements when reading about possible and impossible events. *Journal of Experimental Psychology: Learning, memory, and cognition*, 34, 1001-1010.
- Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavior Research Methods Instruments and Computers*, 20, 6-11.

- Wolff, P., & Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition*, *12*, 153-176.
- Wu, Y. J., & Thierry, G. (2010). Chinese–English Bilinguals Reading English Hear Chinese. *The Journal of Neuroscience*, *30*, 7646-7651.