This study tests the time course of language activation in reading for translation. Reading for translation has been modeled vertically (two monolingual systems activated serially), horizontally (both monolingual systems automatically activated in parallel), and as some blend of these perspectives. Schaeffer, Dragsted, Hvelplund, Balling, and Carl (2016b) provided evidence supporting early horizontal processing in reading for translation. Translators displayed longer first fixations when a word (for example, Spanish *grande*) had been translated in more than one way (*big* and *large* in English). This has a parallel in monolingual studies, where all meanings of polysemous words can automatically be considered or accessed at an early stage (Onifer & Swinney, 1981).

In the context of reading for translation, these findings suggest the study of words that are polysemous in the source language (for example, Spanish *dedo*), but where each meaning has a distinct translation in the target language (*finger* and *toe* in English). The horizontal model predicts automatic early activation of all translations of such words in the target language. On the other hand, the vertical model argues against early activation of target language words.

One experiment, using eye tracking methodology, and one observational study, based on the extensive CRITT database of recordings of different translation tasks in several language pairs, were carried out to investigate the influence of dual language activation on reading for
translation. To study the effects of cross-linguistic polysemy, participants sight translated sentences with ambiguous and non-ambiguous control words in neutral context. The observational analysis completed the experimental data by focusing on an English-Spanish subset of the CRITT database to assess how word translation entropy, a measure of variability of word translations in actual translation output, influenced the length of first fixation durations on a source text. Results at the micro-scale (one language pair, one translation modality) replicated to some extent macro-scale (six language pairs, four translation modalities) results and open the way for further research.
CONTEXT EFFECTS IN READING FOR TRANSLATION:
EARLY TARGET LANGUAGE ACTIVATION

A dissertation submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

by

Anne Frédérique Marie Neveu

August 2018
© Copyright
All rights reserved
Except for previously published material
Dissertation written by
Anne Frédérique Marie Neveu
B.A., ISIT - Institute of Intercultural Management and Communication, 2011
Ph.D., Kent State University, 2018

Approved by

_______________________________, Chair, Doctoral Dissertation Committee
Isabel Lacruz

_______________________________, Members, Doctoral Dissertation Committee
Erik Angelone

_______________________________
Greg M. Shreve

_______________________________
Ryan T. Miller

_______________________________
William E. Merriman

Accepted by

_______________________________, Chair, Department of Modern & Classical Language Studies
Keiran J. Dunne

_______________________________, Dean, College of Arts and Sciences
James L. Blank
# TABLE OF CONTENTS

TABLE OF CONTENTS ................................................................................................. v

LIST OF TABLES ......................................................................................................... ix

LIST OF ABBREVIATIONS ......................................................................................... x

DEDICATION ................................................................................................................ xi

ACKNOWLEDGEMENTS ............................................................................................ xii

CHAPTER 1 .................................................................................................................. 1

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

Translation Studies: Scope and Development ............................................................ 1

Translation Process Research ..................................................................................... 3

Methodologies ........................................................................................................... 5

Think-Aloud Protocols ............................................................................................... 6

Limitations and Developments of TAPs ................................................................. 7

Keystroke Logging ..................................................................................................... 8

Eye Tracking ............................................................................................................... 10

Triangulation ............................................................................................................. 12

Reaction Time Measures ......................................................................................... 13

Physiological Measures ............................................................................................ 14

Survey Methods ........................................................................................................ 16

Research Problem and Significance ......................................................................... 17

Summary of Chapters ............................................................................................... 19

CHAPTER 2 .................................................................................................................. 20
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implications</td>
<td>92</td>
</tr>
<tr>
<td>Modeling Reading for Translation</td>
<td>92</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>93</td>
</tr>
<tr>
<td>Further Research</td>
<td>94</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>96</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>108</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>110</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>113</td>
</tr>
<tr>
<td>APPENDIX D</td>
<td>117</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Overall \textit{FFDur} Means and Standard Deviations (in Parentheses) on Ambiguous and Non-Ambiguous Words per Participant and L1 in Milliseconds

Table 2. Calculation of \textit{THtra} for the Source Language Word ‘imprisoned’ in BML12

Table A1. Ambiguous Spanish Words and their two Translations, with Frequency Ranking, Length and Mean Values

Table A2. Non-Ambiguous Spanish Words and their Translation, with Frequency Ranking, Length and Mean Values

Table C1. Linguistic Experiences and Self-Assessed Proficiency Ratings of the Translation Student Participants (n=12) Working with English and Spanish

Table D1. Data Collected per Participant as a Mean of all First Fixation Durations (\textit{FFDur}) on Ambiguous and Non-Ambiguous Words in the Sight Translation Task in Milliseconds, Depending on the Native Language (L1)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>translation studies</td>
</tr>
<tr>
<td>TPR</td>
<td>translation process research</td>
</tr>
<tr>
<td>ST</td>
<td>source text</td>
</tr>
<tr>
<td>TT</td>
<td>target text</td>
</tr>
<tr>
<td>SL</td>
<td>source language</td>
</tr>
<tr>
<td>TL</td>
<td>target language</td>
</tr>
<tr>
<td>DTS</td>
<td>descriptive translation studies</td>
</tr>
<tr>
<td>TAP</td>
<td>think-aloud protocol</td>
</tr>
<tr>
<td>AOI</td>
<td>area of interest</td>
</tr>
<tr>
<td>L1</td>
<td>first/native language</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>ERP</td>
<td>event-related potential</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>UAD</td>
<td>user activity data</td>
</tr>
<tr>
<td>RfT</td>
<td>reading for translation</td>
</tr>
<tr>
<td>RHM</td>
<td>Revised Hierarchical Model</td>
</tr>
<tr>
<td>BIA</td>
<td>Bilingual Interactive Activation (model)</td>
</tr>
<tr>
<td>CI</td>
<td>Construction-Integration (model)</td>
</tr>
</tbody>
</table>
I dedicate this dissertation to my parents, Marie-Christine and Luc, for their unconditional love and support, and to my brother, Marc, and my sister-in-law, Svetlana, for sharing their experience in the Ph.D. journey.
ACKNOWLEDGEMENTS

I first would like to thank my advisor, Dr. Isabel Lacruz, for her encouragement and guidance throughout the dissertation process. Her outstanding mentorship and support has been instrumental in the success of this research.

I would also like to thank the members of my dissertation committee, Professor Gregory Shreve, Dr. Erik Angelone, Dr. Ryan Miller and Professor William Merriman, for their interest and investment in my project. Their expertise and feedback allowed enhancing the quality of this dissertation.

The norming studies and the Experiment described in Chapter 3 would not have been possible without the contributions of several people. I would like to thank Dr. Moritz Schaeffer who provided me with some advice on the design of the experimental sentences. My gratitude also goes to the professors who let me in their classrooms to conduct the norming studies: Dr. Indira Sultanič, Dr. Jessie Carduner, and Dr. Kelly Washbourne, and to the students who participated in the norming and the Experiment, and who showed interest in my research.
CHAPTER 1

INTRODUCTION

Translation Studies: Scope and Development

The word translation is rooted in the Latin term ‘translatio,’ which encompasses the meanings of ‘moving a thing from one place to another’ and ‘the transference of ideas from one context/language to another’.\(^1\) Translation in its current sense includes written translation and oral or spoken translation, referred to as ‘interpreting’ or ‘interpretation.’ Written or spoken translation can designate both the translation product (or target text –TT–, translated from a source text –ST–) and the translation process (the act of translating).

Whereas the act of translating has a long-standing history, the field of Translation Studies was created in the twentieth century. In a seminal paper originally presented in 1972, James S. Holmes identified the need to create a new field, as opposed to a branch of Linguistics or Comparative Literature among other fields involving translation, because “the paradigms or models fail to produce sufficient results and researchers become aware that new methods are needed to approach the problem” (Holmes, 1991, p. 31). Holmes argued that the field cannot  

grow if its name is not agreed upon and if its scope and structure are not well defined. He was thus the first scholar to delineate a field for the study of translation, which he termed “Translation Studies” (henceforth abbreviated as TS). He determined the focus of the fledgling field as “clustered around the phenomenon of translating and translations” (Holmes, 1991, p. 31) and thus posited that TS is an empirical discipline (Holmes, 1991, p. 35); therefore, it lends itself to experimental or observational study.

Holmes proposed a map of TS dividing it into “pure” TS and applied TS, which, along with their subdivisions discussed below, mutually inform and contribute to each other. Applied TS concerns the practical uses of translation, namely translator training, translation aids and translation criticism. “Pure” TS is a meta-reflection on translation and consists of two main branches for each of the field’s main objectives: Descriptive Translation Studies (DTS), “to describe the phenomena of translating and translation(s) as they manifest themselves in the world of our experience” (Holmes, 1991, p. 36) and Theoretical Translation Studies (ThTS), “to establish general principles by means of which these phenomena can be explained and predicted” (Holmes, 1991, p. 36). Theoretical TS is subdivided into general translation theory and partial translation theories. The aim of a general translation theory is to “develop a full, inclusive theory accommodating so many elements that it can serve to explain and predict all phenomena falling within the terrain of translating and translation, to the exclusion of all phenomena falling outside it” (p. 38). Partial translation theories instead focus on specific aspects of translations, such as the medium through which they are produced (e.g.: human or machine translation), the languages and cultures they involve, the level of textual study (micro- or macro-textual), the text or discourse types and what specific issues they incur, the time at which translations were produced and what problems are typically encountered in translation.
The other branch, DTS, is subdivided into product-oriented, function-oriented and process-oriented DTS. While product-oriented DTS aims at describing existing translations, function-oriented DTS discusses what role(s) translations play in the recipient culture. Process-oriented DTS concentrates on the translation task to answer the following question: “What happens in the translator’s mind while translating?” Although Holmes emphasized the interplay between each branch, he argued that TS, as an empirical discipline, may be most informed by DTS “as the branch of the discipline which constantly maintains the closest contact with the empirical phenomena” (p. 36). As a result, Holmes initiated a move away from solely studying translation products in a deductive fashion to empirically studying translation processes in order to further understand the product. This dissertation is a contribution to translation process research. In the next section we give a brief overview of its scope and development.

Translation Process Research

The study of the translation process fits into a psycholinguistic model of translation (Neubert & Shreve, 1992, p. 29-30) that aims to answer the question formulated above, in the words of Krings (1986): “What goes on in the mind of the translator?” Toury (1995) argued that the translation process cannot be accessed directly, rather it is a “black box” that the researcher can tap into to uncover the cognitive processes at play during the translation process. This formed the onset of Translation Process Research (TPR), a new “empirical, experimental, descriptive approach to translation studies based on close, technology-supported observation of translational (micro)behavior” (Jakobsen, 2015). Moreover, according to Saldanha and O’Brien (2014, p. 109), research on translation processes tries to understand “translator or interpreter
behavior, competence, expertise, the cognitive processes that orient these [translation processes] and the relations between cognition and the translated or interpreted product.”

Twenty years ago, the volume *Cognitive Processes in Translation and Interpreting* (Danks, Shreve, Fountain, & McBeath, 1997) was the first attempt to bring together researchers from the field of Translation Studies and Psychology. This volume set the stage for research into the cognitive aspects of translation by arguing for

“(…) interdisciplinary collaboration, theoretical rigor, methodological innovation, and, above all, for the integration of concepts and rudimentary models of the translation process research of the time with new models of bilingualism, second language acquisition, memory, and cognition that were emerging in psycholinguistics and the other cognitive sciences”

as explained in the subsequent volume, *Translation and Cognition*, published more than ten years later, in 2010 (Shreve & Angelone, 2010a, p. 1). This latter volume was published to present new work using methodological innovations – notably keystroke logging and eye tracking – and resulting findings since 1997 and was a testimony of the integration of concepts and models from the cognitive sciences into TPR originally called for in *Cognitive Processes in Translation and Interpreting*.

Since then, TPR has kept growing, expanding the discussion on theoretical and methodological progress and contributing to enhancing both the translator’s work environment and translator training, while keeping an eye on the future of the discipline. This discussion has been evidenced in the publication of several volumes including, but not limited to, *Cognitive Explorations of Translation* (O’Brien, 2011), *Descriptive Translation Studies - and Beyond* (Toury, 2012), *The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science* (Schwieter & Ferreira, 2014), *New Directions in*
Empirical Translation Process Research (Carl, Bangalore, & Schaeffer, 2016a), and The Handbook of Translation and Cognition (Schwieter & Ferreira, 2017). 2

In their original volume, Shreve and Danks asked six broad research questions that are still at the core of translation process research today (1997, viii-ix). The present dissertation is a contribution to seeking an answer to a blend of two of these questions: “What is the relationship between bilingual language processing and translation/interpreting skill?” and “What methods and models can be used to investigate the cognitive processes of translation and interpreting?” (Shreve & Danks, 1997, viii-ix). Our specific focus will be on monolingual and bilingual models of language processing in translation process research as will be further explained below and in Chapter 2.

Methodologies

TPR relies on three interdisciplinary pillars: the first is the postulate in cognitive psychology that behavior and mind are correlated, the second corresponds to research on the human brain in neuroscience and the third is the methodological rigor borrowed from cognitive psychology and experimental psycholinguistics (Jakobsen, 2017, p. 23). A significant challenge faced by TPR has been to adapt these frameworks to the translation task, all the while retaining ecological validity, that is remaining as close as possible to real-life translation situations (Jakobsen, 2017, p. 23).

2 See Jakobsen (2017, p. 38) for a comprehensive list of volumes contributing to TPR since 2008.
Think-Aloud Protocols

Let us now retrace the progression of methodological innovations since the inception of TPR. The first empirical data gathered as part of TPR was collected in the 1980s with Think-Aloud Protocols (TAPs), a methodology borrowed from cognitive psychology (Shreve & Angelone, 2010b). TAPs spring from Ericsson and Simon’s hypothesis (1984) that if humans can formulate the contents of what is currently and consciously held in mind, more precisely in working memory, then they can report on it, either simultaneously to a task, or retrospectively. The oral data gathered is recorded and later transcribed into written reports or protocols, which can be coded for further analysis (Gerloff, 1986). TAPs provide information as to translators’ strategies (Gerloff, 1986; Krings, 1986), their evaluations of potential ST equivalents, their performance (Laukkanen, 1996), their involvement in the task (e.g.: precision versus vagueness in their reports, first-person references) (Jääskeläinen, 2000), the length of their units of analysis (from the morpheme or syllable to the sentence level) (Gerloff, 1986) and their overall approach to the translation task (Krings’ translation maxims [1986], e.g.: does the translator read the full ST before translating, are problems highlighted, is the audience considered, etc.?).³

Moreover, TAPs provide an understanding of the different levels of processing at different stages of expertise by looking at automatized processes: students learning translation seem to show more processing compared with professionals or non-professionals (Jääskeläinen, 2000). TAPs also allow tapping into the affective factors involved in the translation process. For example, a high level of involvement in the task, self-confidence and a relaxed environment are positively correlated with higher quality performance (Jääskeläinen, 2000; Kussmaul, 1995;

³ See Bernardini (2001) for a discussion of these indicators in TAPs.
Laukkanen, 1996; Tirkkonen Condit & Laukkanen, 1996). TAPs can be performed as a monologue (solo), or as a dialogue between translators. There might be an advantage of dialogue over solo TAPs in that they favor natural discussion and help gain richer reports on the process: it has been shown that the more translators gain in experience, the fewer thoughts they elicit during TAPs (Kussmaul, 1995, p. 91), however, research is not conclusive on which method yields the most useful results (Jääskeläinen, 2000, p. 77-8). Therefore, sharing on the task with a fellow translator helps access processes that may not have been formulated in a monologue TAP (Jääskeläinen, 2000, p. 77-8).

**Limitations and Developments of TAPs**

While TAPs made it possible to access the translation process for the first time, this method has limitations. Indeed, concurrent verbalization can slow down the process, which can be negative in that the reports will be based on a task that does not reflect a real task in terms of process rhythm; however, the slowing down of the process can foster metacognitive reflection on the process, thus providing richer reports (Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995). The main criticism formulated against TAPs is that they disturb the normal course of the translation process, thus leading to skewed verbal reports (Toury, 1995). This criticism was qualified by Jääskeläinen (2000), who argued that the level of distraction incurred by TAPs on the translation process was overestimated, which was supported by Tirkkonen-Condit (2000), especially for concurrent TAPs versus retrospective TAPs.

Another criticism is based on the fact that protocols only show well-ordered thoughts and are thus unable to tap into random thoughts in working memory (Hönig, 1992; Kussmaul &
Moreover, protocols tend to reflect the rationalized thoughts of the participants, which results in the skewing of the contents of the process (Kussmaul & Tirkkonen-Condit, 1995, p. 183): participants share what they thought they did, as opposed to what they actually did. This is also the case in retrospective protocols, when participants are asked to verbalize what they did after the task. Finally, to allow replication and generalization of results, research designs need to be fully explained and need to involve more participants, sampled from a more representative pool beyond that of convenience sampling consisting of translation or language students available at the researcher’s institution (Bernardini, 2001; Shreve & Angelone, 2010a, p. 5). While TAPs provide access to conscious parts of the translation process, unconscious processes remain unknown.

Keystroke Logging

By the end of the 1990s, scholars began to use innovative and more precise methodologies to capture translation process data. Keystroke logging is one of these methodologies, which can record all computer keystrokes during a translation task, and Translog is one of the software applications for keystroke logging that is widely used in translation process research (Jakobsen & Schou, 1999). Translog’s interface is a split window with the source text (ST) in the upper part and the target text (TT) being typed by the translator in the lower part. This software provides data in the form of a screen recording of the Translog window that can be replayed to analyze additions, deletions, pauses, typos, mouse clicks and cursor movements while producing the TT (Carl, 2012). These typing events notably provide information on translator uncertainty and in turn, on monitoring in translation, to answer such
questions as: “How do translators manage uncertainty depending on their level of experience?” and “What are strategies to overcome uncertainty?” (Angelone, 2010, p. 18). Uncertainty is a form of processing difficulty as it is connected to problem recognition and solving (Angelone, 2010, p. 17). Pauses are also evidence of processing difficulty, in monolingual processing (Schilperoord, 1996), translation and interpreting (Timarová, Dragsted, & Hansen, 2011), sight translation (Shreve, Lacruz, & Angelone, 2011) and post-editing (Lacruz, Shreve, & Angelone, 2012); they are markers of increased cognitive effort. For example, in post-editing, keystroke logging allows to obtain measures of time spent per word and per segment and to measure the length of pauses during the process. Two measures of cognitive effort in post-editing emerged as a result of the observed behavior of the post editor facing difficult edits: the average pause ratio, and the pause-word ratio (Lacruz & Shreve, 2014). The latter measure will be further discussed in Chapter 3.

An advantage of keystroke logging over TAPs is that the translation process is not disturbed: keystrokes are logged in the background without interfering with or interrupting the writing process (Carl et al., 2016a, p. 5). In addition, the data gathered shows exactly what happened in the production process, removing any concerns over accuracy, in contrast to TAPs. In addition to showing translators’ typing activity, screen recording provides additional time-stamped data of web or document searches, cursor movements and mouse clicks during the translation process (Angelone, 2010, p. 28), measures that provide further information on the translation process. Screen recording can also serve as a cue during retrospective protocols: subjects are asked to recall and explain what they did while watching the replay of their activity. Cued retrospective protocols help base verbal reports on visible steps of the translation production process, potentially allowing for more thorough and accurate results compared with
Think-Aloud Protocols alone, as screen recording can reveal unconscious processes to the translator. The time lapse between the process and the retrospective recall may however lead to some forgetting of the task process, thereby limiting the effectiveness of cued retrospective protocols.

Eye Tracking

While keystroke logging provides information on the production processes in translation, it does not provide insights into the comprehension processes occurring while reading the ST. That is why eye tracking is another methodology used in TPR, to supplement keystroke logging or used on its own. Like TAPs, this methodology is borrowed from cognitive psychology: Just and Carpenter (1980) argued that what is fixated by the eyes is what is being processed by the mind. They termed this the eye-mind assumption. They also argued that word interpretation must be completed as soon as possible –this is the immediacy assumption. Rayner and Pollatsek (1989) added to the immediacy assumption by explaining that eye movements are indeed under cognitive control, however not necessarily resulting from words currently fixated but rather from recently acquired information. Further research is needed to understand the delay between what is fixated and what is being mentally processed. For example, the eyes can see to the left and right of the target word fixated. This results in interaction with the information content of the immediately surrounding context of the word being fixated and impacts the processing of the fixated word and the surrounding words. This allows pre-processing of the following word (parafoveal preview), but also continued processing of the preceding word (spillover effect). While these are clear benefits to overall language processing during reading, there can also be
downsides, including potential interference with the processing of the fixated word (White, Rayner, & Liversedge, 2005, p. 892). As a result, eye tracking may not yield consistent data across studies, depending on the immediate environment of the target words or areas of interest. One way to try to address this issue and enhance the possibility of replicating and generalizing results is to describe in detail experimental materials, apparatus, displays and procedures.

There are different types of eye trackers: head-mounted eye trackers, best suited for studies that require the participant to move, and desktop-based eye trackers, best suited for experiments in which the participant will sit in front of a computer monitor (Saldanha & O’Brien, 2014, p. 136). Head-mounted eye trackers come in the form of a helmet with two cameras nearby the eyes to track their movements. Although this apparatus is cumbersome to wear, it is still used for research that requires highly precise measures. Eye tracking glasses are now available, making it possible to investigate real-life behaviors (e.g.: of consumers, drivers, gamers).

In translation process research, desktop-based eye trackers are used because the translation task is normally conducted on a computer. The eye tracker resembles a standard computer monitor, fitted with infrared cameras that reflect the light off the eyes (Saldanha & O’Brien, 2014, p. 136). The eye tracker comes with eye tracking software that can calculate the coordinates of the eye path on the monitor with millisecond (ms) resolution. The researcher can use the eye tracking software to create areas of interest (AOIs) on which the device will record data. The data can be viewed as gaze plots or in tables, showing a variety of measures such as fixation count and duration, number of regressions, total time spent on AOI or pupil size, all indicative of cognitive effort. Cognitive effort is defined as the use of mental or cognitive resources, such as working memory, which temporarily stores necessary information for a given
task at hand. More and longer fixations, more regressions, longer times spent on AOIs and pupil dilation and constriction are indicative of higher cognitive effort (Saldanha & O’Brien, 2014, p. 136-142).

**Triangulation**

The eye tracker can be used in combination with software such as Translog II to compute correlations between eye movements and typing behavior (Carl, 2012). This type of combination is called triangulation. It implies “the use of two or more data acquisition methodologies within a single study to improve the quality, validity and reliability of research findings” (Shreve & Angelone, 2010a, p. 6). The eye tracking data provides information on visual attention on the ST and TT, back and forth transitions between comprehension (reading) and written or oral production in respectively written or sight translation, and the time taken between reading a given passage and typing or speaking its translation (Dragsted, 2010, p. 42; Shreve, Lacruz, & Angelone, 2010, p. 63). The keystroke logging data provides information about the production of the target text. This combination of empirical data allows further understanding the coordination of comprehension and production processes or reading and writing/speaking processes in translation. Eye tracking data alone shows unconscious processes ongoing during reading for translation. However, eye tracking data can be difficult to gather due to the sensitivity of calibration. For example, if a subject’s gaze is focused outside of the screen, the initial calibration completed at the beginning of the experiment is lost, resulting in loss of data. Eye tracking data alone does not allow access to production processes. Thus, there is great advantage to using it in combination with keystroke logging and/or TAPs to provide more complete
information. This does not apply to studies focusing on oral translation or only on the reading process in translation, as they do not involve the typing of the translation.

Reaction Time Measures

While eye tracking data helps to further understand the coordination of reading and writing in translation, reaction time experiments enable researchers to investigate the coordination of the first and second language (L1 and L2) when used in translation. The first language is understood as being either the translator’s native language, or the language that is mastered at the highest level, while the second language is typically a language learned after the native language but in which the translator has achieved mastery for translation purposes, at least in reading and writing.

Reaction time tasks are borrowed from psychology. For example, they are used in psycholinguistic studies on the bilingual lexicon, which include lexical decision tasks, translation verification tasks, semantic priming, sentence priming, picture naming, word association, semantic categorization and the Stroop interference task (Pavlenko, 2009, p. 126). The data gathered from reaction time experiments is in the form of time stamps in milliseconds, and results can be statistically analyzed. These tasks make it possible to examine “whether L1 and L2 words in the bilingual lexicon are linked to a common conceptual representation” (Pavlenko, 2009, p. 127). Faster, compared with slower reaction times tend to show stronger connections between L1 and L2 suggesting shared meanings across languages (Pavlenko, 2009, p. 127). For example, De Groot (1992) studied the role conceptual representations play in bilingual lexical representations, focusing on the effects that different word characteristics such as word
imageability (the extent to which a mental image is easily formed), word frequency and cognate status have on bilingual language processing, measured by reaction times (called ‘response times’ in this article). For a review of reaction time studies that assess how meaning is represented in the mind of bilinguals, see Kroll (1993).

Recent studies combine reaction time measures with other measures such as Event-Related Potentials (ERPs) – further discussed below – as in Bice and Kroll’s study of changes occurring in L1 when learning an L2 (2015). The advantage of the reaction time methods is that they can show online language processing, providing a window into the interplay between languages in the mind of the bilingual. However, reaction time experiments are limited in that faster processing times may not be only affected by shared meanings but also by individual differences in L1 and L2 proficiency, time and context of language acquisition, context of habitual use, and word specificities such as frequency of co-activation of word pairs and word similarity (De Groot, 1995; Kroll & Tokowicz, 2005; Marian, 2009, p. 127). In sum, reaction times show a connection between a bilingual’s languages but do not show the extent of that connection (Pavlenko, 2009, p. 128).

**Physiological Measures**

Physiological measures are additional methodologies borrowed from psychology and neuroscience (Saldanha & O’Brien, 2014, p. 148). Electroencephalography (EEG) measures brain activity in response to a stimulus through electric variations resulting from the activity of neuron assemblies (Saldanha & O’Brien, 2014, p. 148) ERPs, mentioned above, are a specific application of EEG as “online measures of brain activity that is directly (and temporally) linked
to cognitive processes (…)’ which ‘(…) allow the various properties of linguistic stimuli to be indexed’ (Meuter, 2009, p. 5). ERPs are primarily used in neuroscience and refer to the positive or negative electric activity (potentials) in the brain, stemming from a stimulus (event) (Jakobsen, 2017, p. 22-3). ERP data is reported in microvolts and can be shown in waveforms (graphs) and scalp topography. Results can also be statistically analyzed. According to Bice and Kroll (2015, p. 2), ERPs “are a sensitive tool for investigating the earliest time course of processing […] they have been used to detect evidence of learning prior to behavioral measures.” As a result, ERPs are useful for investigating early online reading processes in translation but need to be combined with other methods if other aspects of the process are to be studied.

Other physiological measures include galvanic skin response, to measure the flow of electricity through the skin, which is positively correlated with increasing perspiration resulting from emotion or increasing cognitive effort (Saldanha & O’Brien, 2014, p. 148). Functional Magnetic Resonance Imaging (fMRI) measures blood flow and oxygenation in the brain (Saldanha & O’Brien, 2014, p. 148), allowing the identification of areas of the brain that “light up” in language tasks. Physiological measures provide direct evidence on brain and body reactions during the translation task, however they need to be combined with user activity data (UAD), notably from keystroke logging and eye tracking, to help explain the effects gathered by physiological data.

4 Two ERP measures are used in language processing research: the N400 and the P600. The N400 shows a larger peak of activity 400 ms after the onset of stimuli that violates semantic expectations (Meuter, 2009, p. 5). The N400 is also larger for words with lower frequency and is reduced with semantic priming (Meuter, 2009, p. 5). The P600 is a positive wave with maximum amplitude after 600 ms, after the onset of morpho-syntactic violations, as well as gender, number and case violations (see Rossi, Kroll, & Dussias, 2014, p.4 for original research references). Recent studies also indicate that the P600 is linked to generally higher processing demands at sentence level (Meuter, 2009, p. 5).
Survey Methods

In addition to experimental methodologies, TPR can also make use of survey methods to gather data. Questionnaires are used in TPR to gather quantifiable information that can be statistically processed to answer research questions. Questionnaires can be used for descriptive, correlational or experimental research designs, making them a versatile tool for TPR. The disadvantage of questionnaires is that if participants do not answer all questions, loss of data occurs, and a bias is created between participants who answered questions that others did not. Saldanha and O’Brien (2014, p. 145-147) also mention contextual inquiry, an ethnographic technique which in TPR consists of observing and interviewing participants in their natural work environment. The advantage of this methodology is that it provides insight into how tools are actually used, as opposed to how participants report they were used – often a gap exists between the two (Saldanha & O’Brien, 2014, p. 146); as was also found with Think-Aloud Protocols. Contextual inquiry takes into account the full environment of the translator, fitting the theory of situated, embodied cognition (Clark, 1998, 2010), which considers physical and psychological states instead of solely focusing on cognitive processes. The methodology thus provides data at the translator’s macro-level, but not at the micro-level, in contrast to data provided by keystroke logging and eye tracking.
Research Problem and Significance

The translation process comprises at least three components: reading of the source text, transfer between L1 and L2 and production of the target text. This dissertation focuses on the early stages of the translation process, specifically on the reading of the source text with the intent of translation. The objective is to study the activation of the target language during reading for translation and the extent to which it is activated from the earliest stages of the translation process.

Reading for translation has been modeled vertically (two monolingual language systems activated serially) (Seleskovitch, 1976), horizontally (both monolingual language systems automatically activated in parallel) (De Groot, 1997; Ibáñez, Macizo, & Bajo, 2010; Macizo & Bajo, 2006; Ruiz, Paredes, Macizo, & Bajo, 2008), and as a blend of these perspectives (Schaeffer & Carl, 2013; Tirkkonen-Condit, 2005). Vertical processing takes place when source and target language processing are dissociated. In contrast, horizontal processing occurs when both source and target languages are processed simultaneously. Vertical processing tends to be deliberate, whereas horizontal processing tends to be automatic and unconscious. Schaeffer, Dragsted, Hvelplund, Balling, and Carl (2016b) provided evidence supporting early horizontal processing in reading for translation. They observed that translators tended to display longer first fixations during eye tracking when the fixations were on a word (for example, Spanish reloj) that had been translated in more than one way (watch and clock in English) by different translators.

The work of Schaeffer et al. raises interesting questions that are the subject of this dissertation. Their evidence for early horizontal processing in reading for translation was based
on a correlational analysis of a large corpus of data (CRITT data) obtained from translations of six texts from English into several languages (including Spanish, German, Dutch, Danish, Hindi, Chinese, and Estonian) collected in several translation modalities (from-scratch translation, editing of machine-translated output without access to the source text and post-editing of machine translations). Automatic processing has been demonstrated experimentally in the early stages of monolingual reading, where, for example, all meanings of polysemous words are available at a very early stage (Morris, 2006, p. 382; Onifer & Swinney, 1981). This is developed in the section “Ambiguity Resolution in Monolingual Reading in Chapter 2. In addition, experimental studies show that bilinguals automatically access both languages at an early stage of monolingual reading (Beauvillain & Grainger, 1987; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007). It is natural, therefore, to ask whether early horizontal processing can be demonstrated experimentally in reading for translation. The first part of this dissertation undertakes such an experimental investigation in sight translation from Spanish to English.

The work of Schaeffer et al. conflates data from several language pairs (English into Danish, Spanish, Estonian, Chinese, German and Hindi) and several translation modalities. Structural and other differences between languages and processing differences between modalities raise the question of whether early horizontal processing might depend on the language pair and the modality. In the second part of this dissertation, we examine this question by looking at whether the correlations found by Schaeffer et al. are still present in the English>Spanish CRITT data in the written translation modality.

Thus, using an experimental and an observational approach, the present work focuses on early activation of the target language in reading for translation. Combining these methods to assess early activation of the target language in translation has not previously been attempted.
This dissertation contributes to the understanding of early stages of reading for translation, in particular the extent to which horizontal and vertical processes are integrated, which will be further discussed in Chapter 2.

**Summary of Chapters**

Chapter 2 will discuss the conceptual framework of this study. It will begin with a literature review of research in monolingual reading and bilingual reading, to be followed by a presentation of the models of reading for translation. Chapter 3 will describe the experimental approach to investigating early processes in reading for translation. Chapter 4 will describe the observational approach to investigating early processes in reading for translation. Chapter 5 will conclude this study discussing implications, limitations and further research.
CHAPTER 2

CONCEPTUAL FRAMEWORK

More than half of the world population is bilingual (Marian & Shook, 2012). In addition, the number of translator training programs has kept growing since the institutionalization of translation after World War II. Caminade and Pym (1998) explain that the number of institutions granting a degree in translation and/or interpreting has kept growing since 1960, starting at 49, to 108 in 1980, and at least 250 in 1994. This number is updated to 555 based on the List of Approved Translation and Interpreting Schools compiled by the American Translator’s Association (ATA) in 2018. This boom in the availability of translation and interpreting programs can be explained by globalization and a growing demand of the market for translators. In addition, the U.S. Bureau of Labor Statistics (2018) predicts an 18% growth in job outlook for the 2016-26 decade, a much faster rate than the average 7% growth. This can be explained by the steep increase in non-English speaking individuals in the U.S., which will fuel employment in this sector.

As a result of this growth, it seems more pressing than ever to better understand the translation process, for theoretical, practical and pedagogical purposes. Indeed, learning more about how the mind of the translator works will provide insights into the human language function, and in turn can serve to improve the practice and teaching of translation and the
development of translation theory. The translation process can typically be divided between reading/listening, transfer, and writing/speaking processes.

The focus of this dissertation is on reading for translation (RfT). RfT has been modeled vertically (two monolingual systems activated serially), horizontally (both monolingual systems automatically activated in parallel), and as some blend of these perspectives. As we will see below in the review of the theoretical models, the evidence relies either on experimental data collected to compare reading for translation with reading comprehension (Macizo & Bajo, 2006; Schaeffer, Paterson, McGowan, White, & Malmkjær, 2014) or on large amounts of data from various tasks and multiple language pairs (7 source languages and 9 target languages [Carl, Schaeffer & Bangalore, 2016b]) assembled within the Center for Research and Innovation in Translation and Translation Technology (CRITT) database. The tasks on which data was collected and gathered in the CRITT database included the following:

- Manually copying (re-typing) a text from the source window into the target window on a computer screen with identical source and target languages [C];
- Monolingual post-editing of machine translation output (without access to the source text) [E];
- Traditional post-editing of machine translation output [P];
- Review of post-edited text [R];
- Translation ‘from-scratch’ [T].

This dissertation aims at complementing existing research on:

- The time course of language activation in reading for translation, and
- The influence of the target language on source language processing in reading for translation.
These questions were studied through experimental and observational approaches. The experimental approach involved one experiment, a sight translation task from Spanish into English, in which eye movements and oral translations were recorded. This is the first time that sight translation has been used in an experiment to study early reading processes in translation. Previously, a study involving sight translation to investigate reading for translation had provided eye movement measures on later reading processes (Alves, Pagano, & Da Silva, 2011, further discussed below). Sight translation involves speaking the translation of sentences as they are read. Compared with written translation, this task gives much less time for reflection before the production of a translation and therefore facilitates access to the very early processes in translation, which begin with reading the source text. As a result, it seems that, compared with other tasks, sight translation would be a task that would well fit the study of early reading processes for translation.

The experiment specifically focused on the effect of target language ambiguity during source text reading. In translation, ambiguity can come from source language polysemy (words with more than one meaning) or from translation output variability, measured by entropy. Schaeffer et al. (2016b) provided evidence supporting early horizontal processing in reading for translation using data from nine studies from the CRITT database, involving copying, from-scratch translating, machine translation post-editing and monolingual editing of machine translation. Translators displayed longer first fixations when a word (for example, Spanish grande) had been translated in more than one way by different translators (big and large in English).

The present study uses an experimental design to provide complementary data. We matched, on length and frequency, words that referred to two different concepts in the target language to non-ambiguous control words (words with only one possible translation). We first created neutral sentence contexts. Then, ambiguous and control words were inserted in the same neutral context sentence so that either interpretation of the ambiguous word was possible and plausible. We predicted that automatic horizontal processes would activate the target language early in the reading for translation process, and we expected this to be evidenced by longer first fixation durations for ambiguous words than for control words.

The second approach we took to studying the time course and influence of target language activation on the source language in translation was through an observational study at the micro-level scale of the ‘from-scratch’ translation data from the English into Spanish language pair in a subset of the CRITT database, the BML12 study. Previously, only macro-level studies had been conducted using the CRITT database, involving various language pairs and tasks as described above. A benefit of a micro-study is that it makes it possible to compare and contrast the experimental data reported in this dissertation (see Chapter 3) with the observational data from the CRITT database (Chapter 4). Another benefit is that it presents the possibility of comparing and contrasting results at the micro- and macro- levels within the CRITT database. This in turn advances our overall objective of contributing to the development and refinement of models of the translation process by extending the current state of knowledge in the reading for translation literature. Therefore the present study, in addition to using the global measure of word translation entropy, $H_{Tra}$, that encompasses all tasks (C, E, P, R, T) in the CRITT database put forward by Carl (2012), proposes a specific measure of entropy for the translation task only: $THTra$. Further details on this measure are given in Chapter 4.
Previous investigations into the existence and locus of automatic horizontal processing in RfT have exploited the potential for semantic ambiguity in interlingual homographs (words with the same spelling but different meanings in the two languages, e.g.: *coin* meaning *money* in English, but *corner* in French) (Beauvillain & Grainger, 1987), the potential for language ambiguity in cognates (words that are spelled the same and share the same meaning across languages, e.g.: *piano* in English and Spanish) (e.g.: Gullifer, Kroll, & Dussias, 2013; Kroll, Dussias, Bice, & Perrotti, 2015; Lauro & Schwartz, 2017; Libben & Titone, 2009; Macizo & Bajo, 2006; Peeters, Dijkstra, & Grainger, 2013), or the potential for production of ambiguity in words varying levels of entropy as assessed post hoc (Schaeffer et al. 2016b).

In monolingual studies, findings showed that all meanings of polysemous words (words with more than one meaning, e.g.: *bank* can designate a financial institution or the side of a river) can automatically be considered or accessed at an early stage. For the experiment reported in this dissertation, this suggested the use of source language words that correspond to two distinct concepts in the target language.

Therefore, to describe the foundations of this study of reading for translation, we will first review monolingual models of the reading process, including the processing of ambiguous words. Because reading for translation involves two languages, we then review bilingual models of reading, including the processing of ambiguous words in two languages. Overall, the monolingual and bilingual literatures both show that meaning selection is non-selective: all the meanings of an ambiguous word are automatically accessed at an early stage in the reading process. Later, we will review models of reading for translation and the predictions they make as to the influence of target language activation on source language processing in reading for translation.
Monolingual Reading Comprehension: The Construction-Integration Model

Among several models of monolingual reading, Kintsch’s Construction-Integration (CI) model was selected because it is one of the most prominent and widely used models of reading comprehension. It is also sufficiently general (McNamara & Magliano, 2009) to apply by extension to reading for translation (Shreve & Lacruz, 2017, p. 132). The CI model fits the translation task in that “translation is source situation model-induced construction of a target situation model” (Shreve & Lacruz, 2017, p. 133).

Kintsch described the basics of the Construction-Integration (CI) model in 1988, and developed it in 1998 to propose a “general processing framework for cognition” (McNamara & Magliano, 2009, p. 306). The CI model is considered the most comprehensive and usable model (McNamara & Magliano, 2009, p. 306-7) and therefore, it has been the most widely applied (Shreve & Lacruz, 2017, p. 132). The CI model is set within a connectionist architecture (McNamara & Magliano, 2009, p. 308; Kintsch, 1998, p. 8), which implies “a network of elementary units, or nodes, each of which has a degree of activation” (Macdonald, 1995, p. 9). Connections between units are weighted, so that depending on the unit level of activation and the weight of connections to other units, these other units will be excited or inhibited. Excitation or inhibition is spread throughout the network of units to eventually excite or inhibit output units. Connectionist models stand in contrast to classical models: the former work in parallel and interactive ways whereas the latter work in serial ways (Macdonald, 1995, p. 9).

The construction element of the model is a step that results from a combination of visual perception of printed words from a text or auditory perception of speech, the reader or listener’s
purpose and his/her background knowledge and experience. Because the focus of this
dissertation is on reading for translation, I will use a reading rather than a listening context to
discuss the CI model.

The initial step of the reading process, visual perception, combines “word identification
in a discourse context” and sentence parsing, sparking potential interpretations, or propositions,
called “idea units” (Kintsch, 1998, p. 4). Both the result of visual perception and the reader’s
goal will determine the knowledge and experience to be retrieved from the reader’s long-term
memory (LTM). Together, these elements form a network of items containing relevant and
irrelevant information for subsequent comprehension (Kintsch, 1998, p. 4). The integration
element of the model corresponds to the selection of relevant versus irrelevant items, functioning
as a constraint-satisfaction system, keeping or activating elements fitting an appropriate
interpretation and discarding, or deactivating non-fitting ones. Because the activation process
will continue until it reaches a stable point of comprehension, that is until a probable
interpretation is reached, this process is called “spreading activation” (Kintsch, 1998, p. 4).

The comprehension process culminates with the episodic text memory (Kintsch, 1998, p.
103), which is a mental representation of the text made of propositions that are activated at
different strength levels depending on their relevance for a given interpretation. Two elements
form episodic text memory: the textbase and the situation model. The textbase is limited to
information derived directly from the text only, such as the structure of the text and its network
of propositions. The situation model corresponds to the inferences made by the reader, a process
that could be described as solving clues posed in the text by way of mobilizing one’s own
knowledge (about the language, the world and the communicative situation [Kintsch, 1998, p.
103] and experience stored in LTM. If the text is highly detailed, the textbase can serve as
situation model. However, a mental text representation will typically be formed of both textual and knowledge-derived information, in different proportions depending on the text’s complexity and the reader’s level of knowledge and experience (Kintsch, 1998, p. 104). The CI process is repeated across propositions and is updated until an entire passage or text is read: this process is thus iterative, and dynamic across reading cycles (McNamara & Magliano, 2009, p. 308-9).

As suggested by Shreve and Lacruz (2017, p. 134), “by factoring models such as the construction-integration model into translation studies, we provide a powerful explanatory apparatus for explicating the notion of transfer.” Indeed, the goal of a translation is to recreate a similar target text situation model to that of the source text for the target audience, by selecting the most relevant mental and textual representations and drawing on linguistic (source and target languages), textual (text form and content in source and target cultures) and audience knowledge (expertise level and cultural expectations). This poses the question of the reading process in another language than the native language, and “the effect of reading in the L2 on text comprehension” (Shreve & Lacruz, 2017, p. 136). Hvelplund (2017) also used the CI model to interpret results from his experiment involving the translation from English into Danish of general news texts: reading for translation yielded more eye fixations than monolingual reading, which was explained by the fact that when reading for translation, more propositions, or interpretations of a text are considered than in monolingual reading to ensure its accurate rendition for the target audience. We will now discuss in detail eye tracking, a method by which reading can be experimentally studied, via a collection of behavioral measures of eye movements during reading.
The Role of Eye Movements in Text Comprehension

To empirically study comprehension processes, Just and Carpenter (1980) measured eye movements and eye fixations, which are the periods when the eyes are stationary between movements. In one experiment, undergraduate students read a series of 15 scientific texts presented in random order and were asked after reading each text to recall its content, without however setting the goal of reading for memorization. Results showed that the more difficult a word, clause or text unit is to process, the more time is spent on that unit. Difficulty was measured by word frequency – lower frequency means higher difficulty, by the amount of integration of information necessary at a given text unit – a larger amount is more difficult to integrate than a lower one, and by inference-making at the end of a sentence – the more inferencing is required, the more difficult a passage is (Just & Carpenter, 1980, p. 329). It follows that gaze duration correlates with the time needed to determine the meaning of a word.

From this experiment, Just and Carpenter derived two assumptions: the immediacy and the eye-mind assumptions. According to the immediacy assumption, the process of word interpretation is undertaken as soon as possible while reading. The eye-mind assumption proposes that the mind is processing the text that the eyes are fixated on. In other terms, words that are fixated are being processed, or identified and assessed for their meaning, and the process is measured by gaze duration, i.e., longer gaze durations are evidence of more difficult processing.

While both assumptions are robust within a theory of reading, Rayner and Pollatsek (1989) curbed the interpretation of the immediacy assumption. Indeed, they argued that although the immediacy assumption posits direct control of the eyes over fixated words while reading, skipped words are not accounted for. Rayner and Pollatsek added the concept of cognitive
control to justify this nuanced approach: eye movements are not only directed by words currently fixated, they can also be directed by recently acquired information.

While we have reviewed a prominent model of monolingual reading that can be adapted to translation, other models of reading comprehension exist which are beyond the scope of this dissertation. McNamara and Magliano (2009, p. 298), who presented a review of monolingual models of reading, point to the fact that “no one model adequately accounts for a wide variety of reading situations that have been observed and the range of comprehension considered thus far in comprehension models is too limited.” Reading for translation is one such reading situation that requires further research within the framework of reading comprehension in one or two languages. It will be further discussed in the section on models of reading for translation.

The models reviewed in this section focus on general and non-ambiguous language processing. We will now turn to the process of ambiguity resolution in monolingual reading to understand why language activation is considered non-selective in monolingual reading. This will serve as a basis for understanding why language activation is non-selective in bilingual reading.

**Ambiguity Resolution in Monolingual Reading**

Studies on lexical ambiguity in monolingual reading, focusing on homographs (words spelled the same but with distinct meanings) and polysemous words (words that have more than one meaning), show that readers rely on contextual clues to disambiguate meaning (e.g.: Duffy, Morris, & Rayner, 1988; Hogaboam & Perfetti, 1975; Onifer & Swinney, 1981). Subject to

---

6 See Eddington & Tokowicz (2015) for a more extensive review of this literature.
debate is the time course of contextual influence on lexical access. Two models were originally proposed, one based on a modular approach to human cognition (Fodor, 1983), and the other, within a connectionist paradigm, on an interactive one (Rumelhart & McClelland, 1982; Morton, 1969). Rumelhart and McClelland (1982) proposed a model of visual word recognition, in which visual input consists of three levels: the feature level, the letter level and the word level, each containing nodes respectively for elements forming the letters, letters and their position in a word, and words. Nodes are activated depending on the potential for the input to contain a corresponding unit. When activation exceeds a certain threshold, it becomes sufficiently strong to activate other nodes, which are likely to contain units that match the visual input. For example, the letter t is consistent with the word take (Rumelhart & McClelland, 1982, p. 61); if the letter a is activated next, it is consistent with the word take, but inconsistent with the word teach. As a result, combinations with te will be inhibited. If a word is more frequent than another, it is identified faster. Lower thresholds of activation and/or stronger activation potentials can explain this effect. In interactive models, lexical access is attained after inhibitory processes discard inappropriate activations. Identification of the meaning of isolated words is therefore delayed when there is a need to resolve semantic uncertainty or ambiguity. Under interactive models, if prior context is sufficiently constraining, lexical access will be limited to the fitting meaning, making lexical access selective (McClelland & Rumelhart, 1981; Morton, 1969).

In contrast, under modular models, lexical access of ambiguous words is exhaustive (or non-selective), regardless of prior biasing context (Onifer & Swinney, 1981). Hogaboam and Perfetti (1975) nuanced this finding and argued that readers access ambiguous words’ meanings in a terminating order, from the most frequent meaning to the less frequent one. The process of lexical access stops as soon as a contextually appropriate candidate is found. In support of this
proposal, it has been found that biased words (“words that have one interpretation that is much more likely than the other(s)” [Morris, 2006, p. 382]), take longer to process when the context supports the less frequent meaning, compared with the dominant meaning (the meaning that is more likely to occur than the other [Morris, 2006, p. 382]), or either meaning of a balanced ambiguous word (“words with relatively equally likely interpretations” [Morris, 2006, p. 382]), or control words. Morris (2006) defines the concept of likelihood as “the probability that a particular meaning is given as the first associative response to the word presented in isolation.”

Eye movement studies (Dopkins, Morris, & Rayner, 1992; Duffy et al., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner, Pacht, & Duffy, 1994; Sereno, Pacht, & Rayner, 1992) showed that balanced ambiguous words in neutral context are the most effortful to process compared with biased ambiguous words (e.g.: bank as a financial institution or the side of a river), or unambiguous words (e.g.: edge). When prior context disambiguates the less likely interpretation of a biased ambiguous word, that is, the less likely interpretation is the correct one, this word becomes more effortful to process than a balanced ambiguous word or a control word under the same conditions (Morris, 2006, p. 383, Pacht & Rayner, 1993).

Therefore, the evidence from these studies suggests that lexical access is exhaustive for ambiguous words embedded in neutral or prior biasing context, supporting the modular models. With regard to ambiguous words in prior biasing context, two models emerge: in selective access models, prior biasing context is sufficient to limit lexical activation only to the fitting meaning in context. In exhaustive access models, prior biasing context does not inhibit the irrelevant meaning in context – unless a word strongly primes the fitting meaning of the ambiguous word in context (Duffy et al., 1988), so that all possible meanings are accessed, leading to longer meaning selection for the fitting meaning in context.
Within the exhaustive model, two views are posited. According to the first one, the autonomous access model (Duffy et al., 1988), prior disambiguating context behaves as neutral context on lexical access, both for biased (or non-equibiased) (Onifer & Swinney, 1981) and balanced (or equibiased) words (Swinney, 1979). According to the second one, the reordered access model, prior disambiguating context affects lexical access differently depending on the ambiguity status – biased or balanced. If the ambiguous word is balanced, prior biasing context will facilitate access to the appropriate meaning compared with the inappropriate meaning (removing the balance between meanings’ likeliness). If the ambiguous word has a dominant and a subordinate meaning, and if the appropriate meaning is the dominant one, it will be accessed faster than the inappropriate meaning. If the fitting meaning is the subordinate one, it is hypothesized that this meaning would be accessed faster than in less biasing context, and that its meaning could be accessed at the same time as the dominant meaning. Results of the Duffy et al. (1988) study with both biased and balanced ambiguous words, embedded in prior and subsequent biasing context (the context disambiguating the ambiguous words respectively comes before the ambiguous word, or after) supports the reordered access model.

It has been shown that monolingual lexical access is exhaustive: all the meanings of polysemous words are accessed before choosing the appropriate meaning, regardless of biasing context or word frequency. This is evidence for non-selective access. Bilingual studies have found similar results with words that are spelled the same across two languages (both languages are activated regardless of the language of the input). This is of relevance to reading for translation as previous studies have shown that the target language is activated early on in translation, as early as the reading process of the source text (Schaeffer et al., 2016b). We will first review major
models of bilingual language processing and then turn to evidence of activation of a bilingual’s
two languages, and discuss the implications for reading for translation.

Models of Bilingual Language Processing

An essential question underlying bilingual language processing models is whether
languages use separate systems or whether languages interact within a global human language
function, using an integrated memory store for words in all languages (Libben & Titone, 2009).
These two views make very different predictions. For example, if languages use separate
memory stores, then language access should be selective, that is a word presented in one
language should not immediately activate the lexical representation of a translation equivalent in
the other language. However, if languages use integrated memory stores, then words that share
form and/or semantics across languages (such as cognates that share form and meaning, and
homographs that share the spelling but have different meanings) should be activated in the other
language when presented in one of a bilingual’s languages, and that would influence processing.
In other words, language access is non-selective for bilinguals. For example, because cognates
would have lexical and semantic representations in bilingual memory, the combination of the
two languages would effectively increase their frequency as compared with a monolingual, and
so facilitates processing. However, when only form is shared but not meaning, as is the case with
interlingual homographs, extra cognitive effort is required to inhibit the inappropriate meaning in
context, which would slow processing down. Most empirical evidence supports a non-selective
view of bilingual language processing. We describe well-developed models that support this
view.
The Revised Hierarchical Model

Translation can be conducted in two directions: from the native language into the second language (L1 > L2) or vice-versa (L2 > L1). This has an impact on source text reading comprehension and transfer processes depending on the translator’s proficiency in the two languages. Any successful model of bilingual language processing must be able to account for the inevitable asymmetries between L1 and L2. The Revised Hierarchical Model (Kroll & Stewart, 1994; Kroll, Van Hell, Tokowicz, & Green, 2010) approaches this by proposing that translation from L1 into L2 requires conceptual mediation, making it a two-step process (linking the L1 word with a concept, and the concept with an L2 word) whereas translation from L2 into L1 only requires word association, which are direct links between L1 and L2 words, making this process quicker than conceptual mediation (but more “superficial”). The model contains an asymmetry in that the switch from the word association route to the conceptual route (a switch that would result from links growing stronger between L1 and L2) takes place gradually as a function of increasing L2 proficiency.

In addition to the issue of directionality in translation, a model of the translation process must take into account the relationships, or shared representations, between words across languages. Malmkjær (2008) formulated the Literal Translation Hypothesis on the assumption that when a direct equivalent of a word exists in the other language, the default mode of translating is literal. If this literal translation eventually does not fit with the context after monitoring the quality of the translation produced, the translation is changed to a better fitting translation in context. Some words have a higher variability potential, either due to source
language polysemy, or due to the variability in possible translations: a word can be translated in several synonymous or even erroneous translations. In any particular data set, the variability of a translation consists of all the translations translators provided for a given word. This is measured by word translation entropy (Schaeffer et al., 2016b) and is denoted by the $H_{Tra}$ measure in the CRITT database discussed above.

**The BIA/BIA+ Model**

The Literal Translation Hypothesis relies on interlingual priming: similar words across languages will prime, or activate each other. The Bilingual Interactive Activation Model (BIA) model (Dijkstra & Van Heuven, 1998; Van Heuven, Dijkstra, & Grainger, 1998), which posits non-selective lexical access in bilinguals, predicts this priming effect. Similarly to the RHM, an L1-L2 imbalance will result from different levels of proficiency in the two languages: typically the L1 activation is stronger than the L2 activation, therefore L1 words are more likely to prime L2 words rather than the reverse.

The BIA model comprises four levels of lexical representation: letter features, letters, words, and language tags or nodes (Dijkstra, 2005, p. 190). Features of the letters (e.g.: a horizontal or a vertical bar) are added to each other to activate letters. The position of letter features allows the activation of compatible letters, and the position of letters allows the activation of compatible words. Words then activate a language tag. In parallel and as the activation progresses, other initially possible letters, words and language candidates are inhibited. Activation of a word is influenced by its frequency and by how familiar the subject is with the word: increased L2 proficiency translates into increased familiarity, faster activation at
each of the four levels and faster subsequent inhibition of unsuitable competitors. The model however only accounts for lexical representations, but not phonological or semantic ones. In addition, the processing of interlingual homographs (false friends) and cognates, as well as context and task effects is not explained in detail. An update of the BIA model, the BIA+ (Dijkstra & Van Heuven, 2002), aimed to fill in these gaps.

The BIA+ model comprises a word identification system that governs the activation of lexical candidates; activation is triggered via a task/decision system (Dijkstra, 2005, p. 197). The identification system takes into account orthographic and phonological representations at sublexical and lexical levels, and comprises language nodes in L1 and L2 as well as a semantic store.

In this model, sentence context (semantic and syntactic) affects the word identification system. However, non-linguistic context, such as effects emerging from task instructions or participant expectations, can only affect the task/decision system and not the word identification system.

**Selectivity of Language Activation in Bilinguals**

Bilingualism implies language interactions: a bilingual is not two monolinguals in one, able to compartmentalize each language (Grosjean, 1989; Kroll et al., 2015, p. 377). These interactions occur at all stages of language proficiencies in a bilingual individual (Degani, Prior, & Tokowicz, 2011; Kroll et al., 2015).

The models described above predict non-selective language activation in bilinguals. This has been confirmed in many empirical studies. For example, Beauvillain and Grainger (1987)
studied how French-English bilinguals processed interlexical homographs by presenting these words in a specific language mode and then by controlling for interlexical homograph frequencies in each language. Results showed that word frequency had more influence than language mode on initial lexical access, supporting non-selective language access. Van Hell and Dijkstra (2002) assessed whether the knowledge of a second and a third language influenced lexical access in the native (and dominant) language, in word association and lexical decision tasks. Results showed that words in the dominant language still activated words in the less dominant (L2 or L3) languages, suggesting that a multilingual’s lexicon is accessed non-selectively. Van Assche, Duyck, Hartsuiker and Diependaele (2009) focused on testing language non-selectivity in bilinguals in a more natural native-language reading situation. They used cognates integrated in sentences and recorded eye-movements. Results showed that cognates were read faster than non-cognates, showing that reading in one’s L1 is affected by the knowledge of an L2.

Overall, these findings suggest that regardless of the level of semantic constraint of the context, bilingual language activation is language non-selective. This means that all meanings of a word in one language primed by a word in the other, and mediated either via orthography or conceptual representations, will be accessed before meaning selection occurs. This selection is achieved via inhibition of irrelevant meanings in context, and the efficiency of this process depends on language proficiency, level of constraint in the context, and word frequency. For example, in the study by Van Hell and Dijkstra (2002) described above, the authors explain that a minimal level of fluency in the non-native (non-target) language is required to let weaker language(s) effects emerge in native language processing.
We will now turn to models of reading for translation to better understand how a translator’s languages interact while initially reading a text for translation. This framework will be put into perspective with models of monolingual and bilingual reading and a discussion of ambiguity resolution in order to formulate predictions on early language activation and target language influence on source language processing in reading for translation.

Models of Reading for Translation

The written translation process is complex and encompasses several steps. It is initiated with source text (ST) reading, which is rarely read in its entirety before target text (TT) writing begins. In sight translation, the expectation is for the translator to start translating as soon as possible during the reading process: this task does not allow long pauses, and therefore the text cannot be read as a whole ahead of translating. Reading for translation involves the use of two languages which, based on the models of bilingual reading reviewed above (the RHM and the BIA+), would be activated non-selectively in reading for translation. We now discuss how reading for translation has been researched from its beginnings in the 1990s up to now.

Reading for Translation as a Special Kind of Reading

Shreve et al. (1993) first explored the idea that reading for translation is potentially different than other types of reading. The hypothesis is that if reading comprehension is part of the translation task, quantitative measures should show how the reading process is influenced by the translation task. Reading for translation, reading for paraphrasing and reading for
comprehension were compared to test whether reading for translation or reading for paraphrasing have a different influence on the reading process than a general comprehension task. Results suggested that this is indeed the case. There was a correlation between problem identification and reading time, and reading for translation took the longest.

Several other experiments have studied how different reading modalities affect translators’ reading speed. The work of Shreve et al. (1993) was extended and supported in an experimental setting by Macizo and Bajo (2004), who investigated how sentence processing in standard reading compares to reading for translation. Sentences were presented word-by-word on a screen and the experiment was self-paced (participants pressed a button when they were ready to see the next word). In one block of sentences, participants had to read, understand and repeat the sentences in Spanish. In the other block, participants had to read, understand and translate the sentences into English. In a second experiment, the tasks were the same, but language directionality was reversed. Results showed that compared with normal reading, online comprehension was slower for translation in both experiments, highlighting that sentence comprehension partly depends on reading purpose and supporting the hypothesis that reading for translation involves horizontal processes.

In Ibáñez, Macizo, and Bajo (2010), language access and language selection in professional translators was further investigated and comparisons were made with bilinguals using cognates and control words from Macizo and Bajo (2006). These words were inserted in sentences to compare reading for repetition and simple reading (without any subsequent task). In two experiments, the authors aimed to answer the following research questions:

- How do translators cope with the cost of the continuous language switches while translating?
What is the mechanism controlling language access during translation that permits both efficient language selection in one language and rapid between-language alternations?

Overall, results showed that lexical processing depended on the experience of participants in professional translation as well as task demands (reading only versus reading and repeating). Indeed, experienced translators were faster at processing cognates compared with control words in the reading for repetition task, showing that they maintained their two languages active. Bilinguals however processed cognates and controls at the same rate, suggesting that only the language of the trial was activated. In the read-only task, all participants processed cognates faster regardless of their translation training or lack thereof. This suggests that when task demands are lower, both bilinguals and translators can keep both their languages activated during sentence reading. In addition, bilinguals were faster than translators in reading cognates in the read-and-repeat task, but not in the read-only task. This finding supports the hypothesis that translators use language co-activation contrary to bilinguals who seem to be using inhibition. This confers an advantage to translators when switching languages, but slows their reading times in more cognitively demanding tasks (e.g.: reading for repetition).

Focusing on different translation tasks compared with reading for comprehension, Jakobsen and Jensen (2008) used knowledge on eye movements while reading at word level to try to understand how the reading purpose influences the reading process and how other language activities such as oral and written translation might affect the reading process. Two reading tasks were conducted (reading for comprehension and reading for translation; neither task was followed by a comprehension test or an actual translation task) and two translation tasks: sight translation and written translation. Data was collected on task time, fixation duration and count, total gaze time with regressions and transitions across areas of interest (AOIs) on ST
and TT. Participants were six students and six professionals. Results showed that sight translation and written translation took the longest, which was explained by the need to produce a translation, oral or written. Reading for translation took longer than reading for comprehension, which suggested that pre-translation occurred. These results also suggested that parallel processing occurs in reading for translation unlike in reading for comprehension. The delay is caused by the interplay between the two languages. This parallel processing is further discussed in the next section, in terms of horizontal processing.

Alves, Pagano, and Da Silva (2011) complemented Jakobsen and Jensen’s study (2008). Their study followed a different procedure, exploring reading for comprehension with subsequent questions, orally summarizing a text and sight translating. The tasks were conducted under two conditions: in the first, the rhetorical structures of texts changed but topics were the same. In the second, the texts were on different topics but had the same rhetorical structure. Results from the first condition were compared with Jakobsen and Jensen’s results (2008) and results across the first and second conditions in Alves et al. were compared. Differences were observed in the first case with regard to total task time: overall, participants took longer to complete the reading tasks in Alves et al.’s study compared with Jakobsen and Jensen’s. This could be due to three factors. The first corresponds to potential differences in participants’ levels of expertise and proficiency - although each study involved six professionals and six students. The second is the potential difference in levels of familiarity with the sight translation task: subjects in Alves et al.’s study were not familiar with the task, and familiarity was not assessed in the original study. The third factor revolves around the instructions given to participants: longer total task times in Alves et al.’s study might be explained by the fact that participants were allowed to re-read the texts as many times as they needed, whereas it is possible that
participants were only allowed to read the text once in the original study. Fixation count results also varied across both studies, which could be explained by differences in eye-tracking data filter configurations and software use. This points to the necessity of reporting filter configurations when describing experimental designs to allow for replication and generalization of data. Overall, results across conditions in Alves et al.’s study were similar, providing additional support for the hypothesis that the differences observed in the results between the two studies could be due to different filter settings and software for the analysis of eye tracking data. It was also found that professionals were faster than students and that students spent more time on ST than TT, a result that was also suggested in Jakobsen and Jensen (2008) and explained by the fact that students need more time to comprehend the ST at the expense of revision processes on TT when compared with professionals.

Rather than comparing reading tasks to highlight their specificities, Danks and Griffin (1997) sought to understand how ST comprehension is arrived at in translation. They suggested that the translation and comprehension processes mutually facilitate each other. They argued that the processing of ST and TT in translation depends on task specificities and individual skill. As a result, the level of ST processing depends on the level of understanding required for an acceptable translation and how much the translator can understand the text. The authors also argued that the translator formulates several possibilities for the TT during the comprehension process instead of fully comprehending, and then subsequently starting to translate. This interaction between translation and comprehension would mutually facilitate each process. It suggests the simultaneous activation and use of the two languages in reading for translation and is consistent with the horizontal approach to translation discussed in the next section.
The Vertical and Horizontal Models

De Groot (1997) introduced the dichotomy between vertical and horizontal translation. The vertical model consists of two monolingual systems used separately to understand the ST in one language and write the TT in another following a process of deverbalization (Seleskovitch, 1976). According to the vertical model, languages are activated on a needs-basis, where one language is used to read the source text – the other language is not activated at this time. When starting to translate, the source text language is ‘switched-off’ and replaced by the target text language. Therefore, the vertical model posits sequential activation of language in translation. In the horizontal model, however, both of a translator’s languages are activated at the same time: while one language is used to read the source text, the language into which the text will be translated is already active, ready for the formulation of target text propositions. This model implies that translators have a bilingual lexicon linked by shared representations, which are the ones activated in the translation process, making the cognitive process of translation language-pair specific.

Macizo and Bajo (2006) further studied the vertical and horizontal models of reading for translation. They tested whether lexical entry type (presence of homographs or cognates versus control words) and memory load (longer or shorter distance within the sentence between target word and disambiguating context word) affect reading for translation and reading for repetition differently. Proponents of the vertical model argue that reading for translation is not different from monolingual reading because languages are activated sequentially. On the contrary, proponents of the horizontal model argue that, in translation, the two languages are activated in parallel, affecting monolingual source text reading for translation, therefore making this task
different from monolingual reading for repetition. If lexical entries and syntactic structures have similar effects on the two reading tasks, the vertical model is supported. However, the horizontal model predicts that homographs should slow down reading for translation while cognates should facilitate reading for translation. In addition, reading more complex syntactic structures to translate into the target language should slow down reading for translation compared with reading for repetition. Results showed that only reading for translation was affected by lexical ambiguity and memory load. In addition, when cognates were added at the end of sentences, their processing was facilitated in reading for translation but did not affect processing in reading for repetition. These results suggest that within-language reading (reading for repetition) and reading for translation involve different processes: in reading for translation, some reformulation already occurs during the reading process, supporting the horizontal model of translation.

Ruiz, Paredes, Macizo, and Bajo (2008) replicated Macizo and Bajo (2006) and varied lexical frequency and syntactic congruency between SL and TL. If TL lexical entries are activated while reading in the SL, TL words with high frequency should facilitate processing when reading for translation, showing evidence for horizontal processing. If SL and TL are activated sequentially, following vertical processing, high or low TL word frequency should not affect ST reading for repetition or reading for translation. Results supported the horizontal approach to lexical activation in reading for translation. In addition, when sentences in Spanish matched English syntax, processing was facilitated but only in reading for translation, not in reading for repetition. Because target language properties affect source language processing in reading for translation, this result also supports the horizontal model in reading for translation.

Congruency manipulations also allowed the evaluation of how early horizontal processes begin when reading a sentence for translation. Each sentence in Spanish was congruent or
incongruent with English syntax at three areas of the sentence. The critical areas consisted of adjective/noun associations placed at the beginning or at the end of the sentence, and verb/relative clause associations placed in the middle section of the sentence. Results showed that congruency effects already emerged as the initial part of the sentence was read, thus showing that horizontal processes occur early in the reading process. It must be noted that this effect was not observed in the word frequency manipulation described in the previous paragraph, where two versions of each sentence were created: one with the critical word in the initial part of the sentence and one in the last part of the sentence. This suggests that the time course of lexical and syntactic activations are different, an effect that is also suggested in Duffy et al. (1988, p. 443). This asynchrony is also the basis for the next model of reading for translation now discussed.

The Monitor and Recursive Models

Tirkkonen-Condit (2005) proposed the monitor model, according to which literal translation is the default mode of translation for both beginners and professional translators until a literal rendering becomes impossible, as assessed by a monitor that evaluates the translation against target language standards. This literal translation mode would be universal, and was evidenced in a study on language-specific items in the Corpus of Translated Finnish (Tirkkonen-Condit, 2004). These “unique” items in the language were found to have significantly lower frequencies in translated Finnish than in original Finnish texts. Unique items cannot be translated literally because they do not have a direct equivalent in the other language; instead they are expressed differently, for example by translating a similar source language expression. Honing the monitor is a skill acquired with expertise. Carl and Dragsted (2012) interpreted Tirkkonen-
Condit’s monitor hypothesis for the mitigation of literal translation errors as being a parallel (horizontal) process at play both during comprehension and production processes in translation. In sum, literal translation would be the result of early horizontal processes, and a monitor would assess the adequacy of the target text formulated literally. If the literal translation does not conform to target language standards, vertical processes then occur to focus on monolingual formulation, limiting the influence of the source language. Vertical processes thus are later processes in translation.

To further understand why translation processes tend to initially be horizontal (literal and automatic), Schaeffer and Carl (2013) tested if the translation task primes semantic and syntactical representations across languages (Vandepitte & Hartsuiker, 2011). If so, shared representations across languages should facilitate the translation process. Specifically, the authors explore whether in translation, representations that are shared between languages are more primed than in reading for comprehension. They used verbatim recall (speaking what was remembered of a sentence previously seen during the experiment) as a measure of priming (Meijer & Fox Tree, 2003; Potter & Lombardi, 1998). Results showed that early on, shared representations are activated, leading to regeneration in the target language through horizontal processing. If the regeneration did not conform to TL norms, then monolingual vertical processes occurred to fix the TT. Vertical processes also monitored ST-TT equivalence. Thus, they found evidence that both horizontal and vertical processes occur in translation (De Groot, 1997): these processes are active at the same time but horizontal processes take place earlier than vertical ones in the reading process for translation. Schaeffer and Carl proposed a recursive model of translation combining strategic problem solving processes (vertical) with automated translation
processes (horizontal), and thus revised the literal translation hypothesis as well as the monitor model (Toury, 1995).

Early Processes in Reading for Translation

Schaeffer & Carl (2013) found that shared semantic and syntactic representations for equivalent words across languages lead to priming in translation. In follow-up work, Schaeffer et al. (2014; 2017) used these findings as a starting point for investigating the literal translation hypothesis. In an eye tracking study comparing reading for translation and reading for comprehension, they examined co-activation of the SL and TL in the reading process for translation. Contrary to Macizo and Bajo (2006), they did not use specific linguistic items, i.e.: homographs and cognates. Instead, they used full sentences, presented in one block as opposed to word-by-word, to make their experiment more ecologically valid, in other words, more similar to a real translation task. Rather than creating experimental sentences, they used the Berkeley Aligner (Schaeffer et al., 2014) to align the novel The Old Man and the Sea (Hemingway, 1952) and its German translation (Hemingway, 1976). They manipulated the number of possibilities for word translation: one source word to one translation versus one-to-many. In reading for comprehension and reading for translation, participants all read in English, their L2, and translated into German, their L1, for the reading for translation condition. This allowed the researchers to probe for language co-activation in both conditions. In one block, professional translators were asked to read for comprehension sentences with one-to-one alignments and one-to-many alignments. Comprehension was tested with verification sentences. In another block, professional translators were asked to think of the translation of the sentence and to hit the key
for the next sentence once they had formulated the translation in their minds, but without typing the translation. Total reading time, average fixation duration, fixation count, regression count and progressive saccade amplitude (measuring text integration processes) were recorded.

Results confirmed previous findings that reading purpose has an effect on the reading process and that in a bilingual task such as translation, languages are co-activated. This was evidenced by longer reading times, and more fixations and regressions in reading for translation compared with reading for comprehension. Specifically, these measures of late processing in reading showed that the target text is integrated during source text reading, suggesting that reading for translation requires more vertical integration processes than reading for comprehension. Based on the horizontal and vertical models of reading for translation, this integration is interpreted as a vertical process amidst the horizontal process of dual-language activation. Results from Schaeffer and Carl (2013) also support this integration of horizontal and vertical processes in reading for translation.

However, contrary to previous evidence (Carl & Dragsted, 2012; Schaeffer & Carl, 2013; Tirkkonen-Condit, 2005) that vertical processes occur late in reading for translation, the longer average fixation durations in the Schaeffer et al. (2017) study were interpreted as showing that source language lexical representations are encoded early on during source text reading.

First fixation durations, which measure the earliest processes in reading, were longer in reading for translation than in reading for comprehension for one-to-many alignments. However, durations were similar in the two tasks for one-to-one alignments. The interaction of task and alignments provides additional evidence for language co-activation in translation and for automatic use of shared representations across SL and TL. Moreover, it would seem that when
source and target share representations as in one-to-one alignments, lexical access in reading for translation is similar to reading for comprehension.

Schaeffer et al. (2016b) further investigated the effect of word translation variation on reading for translation using nine different data sets within the CRITT TPR-DB. Studying the amount of target text variation as evidence of processing effort was originally proposed by Campbell (2000) with Choice Network Analysis (CNA): the more variation in the translation of a word, the higher the cognitive effort is applied in the translation process. This corresponds to product-based evidence of cognitive effort; hence it does provide information on the cognitive processes at play when translating words with many or few alternatives. Dragsted (2012) experimentally tested and confirmed the CNA predictions with process-based evidence from eye tracking and keystroke logging measures. Schaeffer et al. (2016b) broke new ground by operationalizing the measure of translation variation, which they call entropy (abbreviated HTra) and by combining it with a measure of syntactic distortion (HCross). They hypothesized that similarly to translation lexical variation, syntactic variation across languages will have an impact on early reading behavior: the more syntactic reordering is needed from a source segment to a target segment, the longer the processing of the source segment is expected to last. This study analyzed the overall effect of entropy and word order on from-scratch translation, post-editing, and target text monolingual editing, therefore taking a global approach to studying early reading processes in translation tasks. Results showed that higher entropy and more syntactic reordering indeed led to longer latencies in eye movement measures. These findings are further detailed in the section on studies conducted within the CRITT TPR database in the next chapter.
In this dissertation, both an experimental and an observational approach were undertaken to study early processes in reading for translation. The experimental approach adds to previous research by comparing the length of first fixation durations on words in one language that refer to two distinct concepts in the other language (these words will be called ‘cross-linguistically ambiguous’ in this dissertation), e.g.: *muñeca* in Spanish can be translated as *doll* or *wrist* in English), versus ‘cross-linguistically non-ambiguous’ control words, or words with only one translation in the target language). The observational approach is so designated because it used already collected data from the CRITT TPR database. The aim was to complement the global data analyses that had previously been conducted with this database (Carl & Schaeffer, 2017; Schaeffer & Carl, 2017; Schaeffer et al., 2016b) and that are further discussed in Chapter 4.

The Present Research

The goal of the proposed experiment was to test early target language activation in translation using words that are ambiguous across the two languages involved. To do so, the processing of cross-linguistically ambiguous words was contrasted with the processing of cross-linguistically non-ambiguous control words in a sight translation task. Sight translation was chosen over written translation because it provides insight into the earliest translation processes while reading a text for translation, without the safety ‘back-up’ of being able to come back to the text. When sight translating, translators must speak the translation of the text they are
reading, without any significant delay - a task resembling simultaneous interpreting of oral speech.

Before, early target language activation had been experimentally tested across tasks (reading for translation and reading for comprehension in Schaeffer & Carl, 2013) with recall as a measure of priming. Early target language activation had also been observed in the CRITT database using first fixation durations to measure the influence of target language activation on source language reading for translation. For example, using data on translation, post-editing and monolingual editing of the target text, in six language pairs, Schaeffer et al. (2016b) found that overall, the more translations a word had and the more syntactic reordering was necessary, the longer the first fixation duration.

In an observational study, this dissertation aims at completing the experimental data presented here with a detailed analysis of a subset of the CRITT database, the BML12 (presented in Chapter 4), focusing solely on the translation task and the Spanish-English language pair (the same language pair used in the experiment). The purpose of this micro-level study is to assess whether results at the macro-level (Carl & Schaeffer, 2017; Schaeffer & Carl, 2017; Schaeffer et al., 2016b) still hold at the single-task and language-pair level, and to compare and contrast experimental results with those from observational research to obtain a more comprehensive view of early processes in reading for translation.
CHAPTER 3

EXPERIMENTAL APPROACH

In the previous chapter, we identified the need to research early reading processes in translation via an experimental method, with a task designed to evidence these processes. The advantage of an experimental design is that a cause and effect relationship can be investigated between independent and dependent variables. In addition, intervening variables can be controlled for, reducing their potential influence and allowing the effects of the manipulated variables to show more clearly. The purpose of the eye tracking Experiment presented here is to test early language co-activation in reading for translation, and thus to test the predictions of various models reviewed in Chapter 2.

Controlled experiments are rare, but needed in Translation Process Research (Schaeffer et al., 2014). Indeed, they permit attribution of observed effects to the manipulated variables and reduce the potential influence of intervening variables. Therefore they can strengthen the foundations of TPR and assist in its development. However, a disadvantage of controlled experiments in TPR is that they do not allow participants to be tested in their normal working environment, which may influence their behavior. In addition, it is not possible for controlled eye tracking experiments to involve typical texts that are translated in real life, because 1) the text has to be displayed in larger font and with greater line spacing than a regular text; 2) the text has to be
short (100-150 words) to avoid scrolling, which would result in losing calibration of the participant’s eyes; and 3) the text has to be manipulated to isolate the variables of interest and control potential intervening variables. We believe that despite these disadvantages, the experimental method is a robust medium to tap more deeply into the translator’s black box, based on the examples set by the experiments in the monolingual and bilingual literature reviewed in Chapter 2.

This chapter is devoted to the experimental part of this dissertation. First, the variables of interest are identified and the experimental hypotheses are presented. Then the preparation for the experiment is explained, followed by a description of the experimental procedure. Finally, results are presented and discussed.

**Hypotheses**

Our purpose is to investigate early language co-activation in reading for translation. We use first fixations on the source text to gain insight into early processing during reading for translation. First fixation duration corresponds to the length of time a word is fixated the first time the eyes encounter that word, and so this measure is adequate to investigate early processes in reading for translation. We anticipate that first fixation durations will be affected by ambiguities across languages (Beauvillain & Grainger, 1987; Van Assche et al., 2009; Van Hell & Dijkstra, 2002) and asymmetries between reading in the first and second languages (Dijkstra & Van Heuven, 1998; Kroll & Stewart, 1994; Van Heuven et al., 1998). We manipulate two independent variables, Cross-Linguistic Ambiguity with two levels (Ambiguous and Non-Ambiguous) and
Reading Directionality with two levels (L1 and L2). We evaluate the effect of these independent variables on the dependent variable First Fixation Duration.

Cross-linguistically ambiguous words are words in one language with translations that refer to two different concepts in the other language. Cross-linguistically non-ambiguous words are words in one language with translations that refer to a single concept in the other language. In bilingual reading, all meanings of cross-linguistically ambiguous words read in neutral context are automatically accessed early in the reading process, as predicted by non-selective theories of bilingual language activation (Dijkstra & Van Heuven, 1998; Kroll & Stewart, 1994; Kroll et al., 2010; Van Heuven et al., 1998). In translation, horizontal, or parallel processing, which is early and automatic, is also based on non-selectivity (Ruiz et al., 2008; Carl and Dragsted, 2012; Schaeffer and Carl, 2013; Schaeffer et al., 2016b; Schaeffer et al.; 2014; Schaeffer et al., 2017). In reading for translation, we therefore anticipate non-selective activation of the meanings of cross-linguistically ambiguous words. If this occurs, we expect that early processing delays will result from competition between the multiple meanings. Accordingly, we hypothesize that First Fixation Duration will be longer in the Ambiguous condition than in the Non-Ambiguous condition.

Reading in L2 is more effortful than reading in L1, as indicated by longer processing times associated with lower familiarity (Dijkstra & Van Heuven, 1998; Kroll & Stewart, 1994; Van Heuven et al., 1998). Accordingly, we hypothesize an effect of Reading Directionality: First Fixation Duration will be longer in the L2 condition than in the L1 condition.

Further, we anticipate an interaction between the two independent variables, Cross-Linguistic Ambiguity and Reading Directionality. We expect that the Cross-Linguistic Ambiguity effect will be more marked for participants reading in their L2, because of the expected compounding of the anticipated Cross-Linguistic Ambiguity effect and Reading Directionality.
effect. We further hypothesize that First Fixation Duration will be longer in the Cross-Linguistic Ambiguity condition for L2 readers than for L1 readers.

To better access early processing in reading for translation, our hypotheses were tested in a sight translation task with Spanish as the source language and English as the target language. The experiment was conducted using sentences where the context is neutral, not biased toward any meaning of the translation of the experimental words in the target language.

If our hypotheses are confirmed, the results of this experiment would indicate that early horizontal processing, previously found in monolingual reading of ambiguous words (Morris, 2006; Onifer & Swinney, 1981) and in bilingual reading (Beauvillain & Grainger, 1987; Van Assche et al., 2009; Van Hell & Dijkstra, 2002), extends to the situation of reading for translation.

Preparation of Experimental Materials

Words and Sentences

The first stage in our preparation for the experiment was the selection of Spanish words that are cross-linguistically ambiguous when the target language is English. These experimental words were selected to control for the following factors:

- Cross-linguistically ambiguous Spanish words were chosen so that each English translation referred to a single concept. For example, a word like *tronco* in Spanish would not be selected. There are two translations into English, namely, *trunk* and *torso*, but while *torso* refers to a single concept, *trunk* can refer to a tree trunk or a car trunk and so is polysemous
in English. Polysemous words require more cognitive effort to process in monolingual settings (Duffy et al., 1988; Hogaboam & Perfetti, 1975; Onifer & Swinney, 1981), so polysemy of a translation could potentially skew results.

- All experimental words and their translations were nouns. It is known that cognitive effort in word processing tends to be lower for nouns than for verbs (Holmes & O’Regan, 1981; Rayner, 1977). Restriction to nouns again reduces the potential for skewed results.

- Experimental words were chosen not to be compound nouns in order to avoid interference that could emerge from the semantics of each part of a compound (Girju, Moldovan, Tatu, & Antohe, 2005).

- Cognates and near-cognates – words that are identical or near identical in form and meaning across English and Spanish (e.g.: piano-piano, acción-action) – and homographs – words that are identical or nearly identical in form but not meaning across English and Spanish (e.g.: red-red) - were avoided because of some uncertainty about whether or not they facilitate or hinder processing (Starr and Rayner, 2001, p.161; De Groot and Nas, 1991; Altarriba, Kambe, Pollatsek, & Rayner, 2001).

Each experimental word was matched to a cross-linguistically non-ambiguous word, which had a single meaning in Spanish and whose translation into English also had a single meaning. We controlled for the following variables as well as possible in the matching process:

- Word length: no more than two letters difference was tolerated between words in a pair. Words in most pairs had the same number of letters or a single letter difference. Controlling for the number of letters is important when interpreting fixation times in reading since longer words take more time to process (Frederiksen & Kroll, 1976; Richardson, 1976,
Weekes, 1997). Therefore, for two words to be compared on an equal basis, they must have similar lengths.

- **Word frequency:** Ideally, each word in a pair would have a similar frequency. This would allow them to be compared on an equal basis, since as word frequency increases, fixation time tends to decrease (Balota & Chumbley, 1984; Frederiksen & Kroll, 1976; Rayner & Duffy, 1986). However, there is no satisfactory Spanish frequency dictionary comparable to widely used English frequency dictionaries such as Francis and Kučera (1982) or more recent publications (Davies & Face, 2006). This meant that it was not possible to match frequencies in a standard way.

Instead, we matched our stimuli as well as possible using frequency rankings of the most frequent words. To increase the likelihood that participants actually knew the words in each pair, the Spanish words were chosen from a frequency ranking dictionary of the 5000 most used words in Spanish (Davies, 2006). We also attempted to match the frequency rankings of the translations, using a frequency ranking dictionary of the 5000 most used words in English (Davies & Gardner, 2010) or the Corpus of Contemporary American English (COCA). Due to the limited supply of reasonably frequent cross-linguistically ambiguous Spanish words with reasonably frequent English translations, it was not always possible to select stimuli so that the frequency rankings of English translations of cross-linguistically ambiguous Spanish words were comparable. On average, the frequency ranking of the less frequent translation was almost twice the ranking of the more frequent translation, and almost half of the less frequent translations were not among the 5000 most used words in English.

---

7 Access at: [https://corpus.byu.edu/coca/](https://corpus.byu.edu/coca/)
Once the cross-linguistically ambiguous words and their matched cross-linguistically non-ambiguous control words were selected, neutral context Spanish carrier sentences were constructed with a blank space. These sentences were to be completed in two ways: a cross-linguistically ambiguous experimental word would be inserted into the blank to create one version, and its matched cross-linguistically non-ambiguous control word would be inserted into the blank to create the other version.

Next, norming studies were conducted to rate the neutrality of the carrier sentences and the familiarity of potential participants with the experimental words and their translations.

**Norming Studies**

The purpose of norming the materials was to ensure that they would be valid tools for measuring the variables involved in our research. Norming has seldom been conducted in reading for translation research, with the exception of Schaeffer et al. (2017), who noticed this issue and encouraged the practice of norming to limit possible effects of intervening variables on experimental results.

A first round of norming was conducted with undergraduate students in a Spanish translation practice course. However, results showed that participants were mostly unable to name alternative translations for the cross-linguistically ambiguous words. Accordingly, a different participant pool was identified, namely, graduate students in a pedagogical grammar course and a Spanish translation course, and the norming studies were re-run.
**First norming study.** Participants completed two paper-and-pencil tasks: a cloze task in which participants filled in the blanks in the Spanish carrier sentences with Spanish words, and a word-in-isolation task, in which they provided all the English translations they knew for the cross-linguistically ambiguous words and their matched cross-linguistically non-ambiguous Spanish control words.

The rationale for the cloze task was to control for the predictability of the experimental words in the supposedly neutral context. If the word was predictable, then its processing would be facilitated, therefore confounding the independent variable of Cross-Linguistic Ambiguity.

The rationale for the word-in-isolation task was two-fold. First, it was needed to ensure that participants could reasonably be expected to know the Spanish words and at least one of its translations into English. Second, it was needed to demonstrate that the participants produced only one English translation for cross-linguistically non-ambiguous Spanish control words, but two translations into English for most cross-linguistically ambiguous Spanish words. In other words, this task served to control for participants’ level of familiarity with the experimental stimuli and their translations into English.

**Participants.** Thirteen participants carried out the norming; the data for one participant had to be discarded because norming instructions were misunderstood. All participants had English as their L1 and Spanish as their L2. The norming study took place during participants’ undergraduate Spanish translation practice class, except for one participant who was contacted directly and did the norming study in the researcher’s office.
**Materials.** For the cloze task, the carrier sentences with unfilled blanks were randomly divided into two equal length lists, A and B. For the word-in-isolation task, a list was created of all experimental Spanish words in random order.

**Procedure.** Participants were first given the cloze task and then given the word-in-isolation task, so that the latter would not prime how they would fill in the blanks in the sentences. As the norming study was carried out during a class, it was necessary to restrict the amount of time it would take to complete it. Accordingly, for the cloze task, half of the participants received list A and the other half received list B. Participants were instructed to fill in the blanks with the first Spanish word that came to mind. The task was self-paced and was expected to last an average of ten minutes. The word-in-isolation task was given to participants as soon as they finished the cloze task. Participants were instructed to provide all the English translations they knew for the Spanish words in the list and were told that some words would only have one translation, while others would have more than one. This second part of the norming study was also expected to last approximately 10 minutes.

**Results.** When four or more participants produced an experimental word in the cloze task, the carrier sentence was assessed to be too predictable of the experimental word. None of the experimental words were predicted; therefore, the sentences were kept as they were. However, only a small minority of the undergraduate participants could provide both the anticipated translations for most of the ambiguous words. Therefore, it was deemed necessary to abandon this participant pool and identify a new pool.
**Second norming study.** The new participant pool consisted of graduate students. In total, 20 graduate students in a Spanish program were recruited for the second norming study, of which 13 were enrolled in a Pedagogical Grammar course for teaching Spanish and 7 were enrolled in a Scientific, Technical and Medical Translation class. The materials and the procedure were the same as in the first norming study. The data from a student whose L1 was neither English nor Spanish was removed, providing a total of 19 results.

**Results.** The cloze task criterion for rejecting an experimental sentence was the same as in the first norming study. No experimental sentences had to be rejected. In fact, no experimental word was predicted by more than two participants.

The results of the word-in-isolation task were better than in the first study. However, while all participants were able to name one English translation for most of the cross-linguistically ambiguous Spanish words, a second translation of many of the Spanish words was named with much less reliability.

Nevertheless, since no other viable participant pools were available, we proceeded with the Experiment, using the most satisfactory words from the norming study and selecting participants from a pool of graduate students in Spanish. The experimental stimuli selected from the norming materials are listed together with their relevant characteristics in Appendix A, Tables A1 and A2.
Experiment

The aim of the Experiment was to investigate whether both languages are active at the earliest stages of reading for translation. Specifically, we studied whether there is non-selective activation of all meanings of cross-linguistically ambiguous words. A sight translation task was used to increase the salience of early stage processing, since the opportunity for reflection is reduced due to the time pressure to speak the translation. Early stage target language semantic activation was measured using first fixation durations recorded by an eye tracker. We predicted that cross-linguistically ambiguous words would have longer first fixations than matched cross-linguistically non-ambiguous words. Such a finding would indicate full semantic activation of the target language at an early processing stage. This, in turn, would be consistent with horizontal processing at an early stage of reading for translation. We also predicted that participants reading in their L2 would display longer first fixations than participants reading in their L1, due to the greater difficulty of the task caused by lower familiarity with that language. Further, we predicted a compounding of these two effects, so that first fixation durations would be longest for cross-linguistically ambiguous words in L2 reading.

Participants

Fourteen participants took part in this Experiment, but the data of two participants had to be discarded: one due to a task instruction issue, the other because less than a third of the data was collected by the eye-tracker. Most participants were first or second year students from the Spanish track in the M.A. in Translation program at Kent State University, or students in the Ph.D. program...
in Translation Studies; one participant was an alumna of the M.A. in Translation and an instructor of Spanish. Participants’ ages ranged from 21 to 35, half had Spanish as their L1 and half had English as their L1, English-Spanish was their main language pair and most were used to translating from and into each language in different settings (at home, for themselves, at work and at school). Half of the participants had professional translation experience outside of coursework as freelancers, via an internship or as in-house translators for a business; the length of experience ranged from 4 months to 11 years, with a median of one year. Participants were tested in a lab setting with conditions similar to an office and were compensated for their time.

Materials

Twenty cross-linguistically ambiguous Spanish words were selected. The mean length was 5.1 letters for the Spanish words, 5.3 letters for the most frequent English translations, and 6.2 letters for the least frequent English translations. The mean frequency ranking was 2077 for the Spanish words, 2116 for the most frequent English translations, and 3689 for the least frequent English translations.

Twenty matched cross-linguistically non-ambiguous Spanish words were also selected. The mean length was 5.5 letters for the Spanish words, 5.1 for the English translations. The mean frequency ranking was 1702 for the Spanish words and 1707 for their English translations.

Frequency rankings greater than 5000 were recorded as 5000, since they were not available more precisely. As noted before, the word stimuli were not matched as closely as we would have wanted, but the constraints imposed by the Experiment severely limited the supply of alternative stimuli, and we were unable to find closer matches that were familiar to the
potential participant pool. Words that had low frequencies or whose translations had low frequencies were ruled out because it was important for participants to actually know the words and their translations.

A Wilcoxon Signed Rank test for paired data was conducted to compare the ranked frequencies of the cross-linguistically ambiguous words and those of the cross-linguistically non-ambiguous control words. Frequency rankings were not significantly different for cross-linguistically ambiguous words (Mdn = 1980) than for cross-linguistically non-ambiguous words (Mdn = 1508), T = 58, p = .079, r = -.39.

A table with the ambiguous words, their two translations, their matched non-ambiguous word and their translation, and the frequency ranking and length of each word can be found in Appendix A, in Tables A1 and A2.

Forty Spanish sentences were included in the experiment; half were experimental sentences and half were filler sentences. Experimental sentences were constructed with neutral context and in matched pairs, as described in the Words and Sentences section. Ten of the experimental sentences contained cross-linguistically ambiguous stimulus words, and the remaining ten contained matched cross-linguistically non-ambiguous stimulus words. The stimuli were placed in the middle of the sentence to avoid initial and final sentence fixation effects (Rayner, 1977; Warren, White, & Reichle, 2009). Filler sentences were also constructed to have neutral context, and each was matched to an experimental sentence. Filler “stimuli” were selected from the frequency dictionary, with the requirements that they had not already been used in an experimental sentence, and that they were cross-linguistically non-ambiguous. A list of experimental and filler sentences can be found in Appendix B. Two lists of twenty experimental and twenty filler sentences were created so that a participant never saw the same
sentence twice, with a cross-linguistically ambiguous word and the version with its matched cross-linguistically non-ambiguous control. To create those lists, the cross-linguistically ambiguous words were randomized and then matched with their controls. The first ten of the cross-linguistically ambiguous word sentences and the second ten of the cross-linguistically non-ambiguous control word sentences were placed in List A, and the rest of the experimental sentences were included in List B. The presentation of sentences was randomized to avoid any possible order effects.

There was a single Area of Interest (AOI) in each of the experimental sentences, extending fifteen characters to the left and five characters to the right of the center of the experimental word. The location and extent of the AOIs was based on evidence on perceptual span during fixations in reading (McConkie & Rayner, 1975). All AOIs were created with a height of 100 pixels.

Sentences were displayed in black on a white background in 16-point Times New Roman font. They were centered on the screen. If a sentence required more than one line, the lines were double spaced and the line break was made at least one word before or after the AOI. This arrangement matched normal expectations for the appearance of a standard sentence (e.g.: sentences were not displayed in a pyramidal fashion as can be seen in subtitles). It also served to prevent the sentence display from affecting eye-fixation results.
Apparatus

A Tobii TX300 eye-tracker was used to record participants’ eye-movements, with a 23”, 1920 x 1080 widescreen monitor. The sampling rate for gaze position was 60 Hz. The software used to analyze the eye-tracking data was Tobii Studio Professional.

Procedure

When entering the laboratory, participants first signed an informed consent form. The expectations were then described. Participants were instructed to speak the translation of the sentences they saw on the screen with as little delay as possible. After completion of the translation of each sentence, they were to press the space bar that made the current sentence disappear from the screen and revealed the next sentence.

Before the presentation of the experimental sentences, a 5-point calibration was made and then participants spoke the translation of five practice sentences. When the translation of the fifth practice sentence was complete, the word ‘Ready?’ was presented on the screen. The participants pressed the space bar to start the experiment.

Participants had previously been instructed that any questions or concerns should be addressed to the researcher when the ‘Ready’ prompt appeared on the screen. This was to avoid disruption of the data collection once the experiment had begun. Participants who asked questions at this stage were then re-calibrated. When the experiment was over, participants were then asked to fill out a Language History Questionnaire and a Demographic Questionnaire to gain knowledge about their background. This information is summarized in the Participants
A two-way mixed ANOVA was conducted to investigate whether cross-linguistically ambiguous words displayed longer first fixation durations than their cross-linguistically non-ambiguous controls. Further, this ANOVA was used to investigate whether participants reading in their L1 or L2 had an impact on First Fixation Durations when reading for translation, and whether there was an interaction between these variables. There were two independent variables. Cross-Linguistic Ambiguity that was manipulated within-subjects and had two levels (Ambiguous and Non-Ambiguous), and Reading Directionality that was manipulated between-subjects and had two levels (L1 and L2). The dependent variable was First Fixation Duration. Participants whose means were more than three standard deviations above or below the overall condition means were considered outliers. Only one outlier had to be removed.

The main effect of Cross-Linguistic Ambiguity was not significant, $F(1, 9) = .47$, $p = .511$. Thus, mean first fixation duration on cross-linguistically ambiguous stimuli ($M = 239$ ms, $SD = 164$) was not significantly different from that on cross-linguistically non-ambiguous controls ($M = 292$ ms, $SD = 208$). The main effect of Reading Directionality was also not significant, $F(1, 9) = .65$, $p = .440$. Overall, the mean first fixation duration for participants who had English as their L1 ($M = 306$ ms, $SD = 181$, 95% CI [151, 461]) was not significantly different from that of participants who had Spanish as their L1 ($M = 231$ ms, $SD = 189$, 95% CI [90, 373]). Effect Size calculation, using Cohen’s $d$, yielded a small to medium effect size of
A post-hoc power analysis showed there was not enough power to detect the effect of this independent variable, $1 - \beta = .66$ and that a sample size of 16 would be necessary for the Reading Directionality effect to be significant. Additionally, there was no significant interaction between the two independent variables of Cross-Linguistic Ambiguity and Reading Directionality, $F(1, 9) = .81, p = .391$. In sum, there was no significant increase in the Cross-Linguistic Ambiguity effect for L2 readers.

Table 1 provides the summary data; detailed data collected per participant is provided in Table D1 (Appendix D) as a mean of all first fixation durations on ambiguous and control words in the sight translation task, depending on the native language (L1). The raw data shows that on average, first fixation durations were longer on the control words than on the ambiguous words, not supporting the hypothesis initially formulated.

Table 1

*Overall FFDur Means and Standard Deviations (in Parentheses) on Ambiguous and Non-Ambiguous Words per Participant and L1 in Milliseconds*

<table>
<thead>
<tr>
<th></th>
<th>Ambiguous</th>
<th>Non-ambiguous</th>
<th>Overall means</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (Spanish)</td>
<td>177 (118)</td>
<td>285 (240)</td>
<td>231 (189)</td>
</tr>
<tr>
<td>L2 (English)</td>
<td>314 (194)</td>
<td>299 (189)</td>
<td>306 (181)</td>
</tr>
<tr>
<td>Overall means</td>
<td>239 (164)</td>
<td>292 (208)</td>
<td></td>
</tr>
</tbody>
</table>
Discussion

Cross-Linguistic Ambiguity

Contrary to our hypothesis, First Fixation Duration was not significantly different in the Ambiguous and Non-Ambiguous conditions. This finding is not consistent with our expectation that target language translations of cross-linguistically ambiguous words would be accessed non-selectively. This was surprising, considering robust previous evidence for non-selectivity of lexical access in monolingual (e.g.: Dopkins et al., 1992; Duffy et al., 1988; Morris, 2006; Onifer & Swinney, 1981; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner et al., 1994; Sereno et al., 1992) and bilingual (e.g.: Beauvillain & Grainger, 1987; Dijkstra & Van Heuven, 1998; Kroll & Stewart, 1994; Van Assche et al., 2009; Van Hell & Dijkstra, 2002; Van Heuven et al., 1998) studies of the processing of words with multiple meanings. Non-selective access implies that all meanings of a monolingually or cross-linguistically ambiguous word are activated within and across languages, slowing down its processing compared with that for a non-ambiguous word.

On the surface, the lack of effect might be thought to be attributable to low power, stemming from the small number of participants. However, the extremely high standard deviations in the $FFDur$ data make that explanation implausible.

The most plausible factor that might explain the absence of the expected Cross-Linguistic Ambiguity effect is inadequacies in the characteristics of the experimental words. As we noted before, the choice of experimental words was very constrained. The cross-linguistically ambiguous Spanish words were well matched for length with their cross-linguistically non-
ambiguous Spanish control words. However, a significant factor may be the fact that one of the translations of the cross-linguistically ambiguous words was often ranked as much less frequent (much higher-ranking number) than the other. In fact, in the norming study, the second translation was often not named, and so was presumably not familiar to many of the norming participants, who were from a similar pool to the experimental participant pool. Accordingly, many of the experimental participants may have responded to the cross-linguistically ambiguous words as if they were cross-linguistically non-ambiguous. If this were the case, the Cross-Linguistic Ambiguity effect would not be apparent. However, the median frequency ranking for ambiguous words (1980) was lower than that for non-ambiguous words (1508). Although this difference is not statistically significant, it did approach significance ($p = .079$). This means that ambiguous words may have had a tendency to being less frequent than non-ambiguous words. This tendency would favor finding the predicted effect of slower first fixation durations on ambiguous words, which we did not find.

**Language Directionality**

Contrary to our hypothesis, First Fixation Duration was not significantly different in the L1 (participants reading in their L1) and L2 (participants reading in their L2) conditions. This finding was not consistent with our expectation of faster lexical access for participants reading in their L1. This is surprising considering previous robust evidence supporting our hypothesis (Dijkstra & Van Heuven, 1998; Kroll & Stewart, 1994; Van Heuven et al., 1998).

There was however a numerical difference in the expected direction in the mean first fixation times for L1 readers (231 ms) and L2 readers (306 ms), although the very high standard
deviations (189 ms for L1 and 181 ms for L2) dwarf the difference (75 ms) between the means and there was not enough power for this effect to be significant.

**Interaction**

The absence of the expected effects renders the absence of an interaction unsurprising.

The unexpected outcome of our experiment and our discussion of possible causes suggest that it might be fruitful to seek other approaches to understanding the effect of language ambiguities on the early stages of mental processing in reading for translation. To achieve this goal, we turned to the CRITT Translation Process research database. This assembles several hundred hours of translation task measures recorded across a larger group of participants than in the present Experiment.

Our goal was to again assess the time-course of dual-language activation in translation through analysis of the first fixation duration ($FFDur$) for cross-linguistically ambiguous words, measured during eye tracking in a from-scratch translation task. We focused on the same English-Spanish language pair, but, in order to be able to make use of the CRITT database, we studied translation in the opposite order, from English into Spanish. This time we took a correlational approach, using a scale for cross-linguistic ambiguity, based on the translation entropy measure $HTra$, which we describe in detail in the next Chapter.
Models of reading for translation discussed in Chapter 2 predicted that cross-linguistically ambiguous words would manifest longer first fixation durations than cross-linguistically non-ambiguous words. However, in the Experiment described in Chapter 3, we found no significant effect of cross-linguistic ambiguity on first fixation duration. As we noted in the discussion of the results of the Experiment, the constraints on the choice of experimental materials may have led them to have characteristics that masked the expected effect of cross-linguistic ambiguity on early eye movements.

In this Chapter, we further investigate cross-linguistic ambiguity in reading for translation, taking a post-hoc perspective and expanding the concept of cross-linguistic ambiguity to be less constraining. Indeed, the cross-linguistic ambiguity discussed in Chapter 3 is rooted in the source language: the semantic ambiguity of polysemous words in the source language yielded two distinct translations in the target language (e.g.: *muñeca* in Spanish can be translated as *doll* or *wrist* in English). Another type of ambiguity in translation is rooted in the target language: variability in word translation can result not just from polysemy in the source language, but also from the availability of synonyms in the target language. It therefore becomes interesting to consider the range of possible translations of a word in an ecological, rather than
purely theoretical, context. This is best studied through the examination of translations of several texts by several individuals.

This is precisely what the Center for Research and Innovation in Translation and Translation Technology (CRITT) offers with their Translation Process Research (TPR) database (DB). The CRITT TPR-DB, further presented below, gathers several studies in one place, with various language pairs, various tasks or translation modalities studied and several measures of eye-movement and typing activity recorded. Target language variation in word translation is measured through the concept of entropy from Shannon’s formulation of Information Theory (1948). The metric used in the CRITT TPR-DB is called word translation entropy and abbreviated $HTra$ (Carl et al., 2016b). In the Data Preparation section below, we describe how this is operationalized.

Previous macro-level research (Schaeffer et al., 2016b) revealed a significant but weak global correlation between word translation entropy, $HTra$, and (the logarithm of) first fixation duration, $FFDur$, in eye tracking studies, with data aggregated across several language pairs, including English into Danish, Spanish, Estonian, Chinese, German and Hindi, and across several translation modalities, including from-scratch translation, post-editing of machine translation (MT) output and monolingual editing of post-editing of MT output. This correlation indicates an association between cross-linguistic ambiguity (measured differently than in Chapter 3) and first fixation durations. It provides evidence of early target language activation during the process of a translation task, and so suggests early horizontal processing. However, the broadness of the sweep of this analysis raises several questions related to the language pairs and the translation modalities:
• Does the global correlation between \( H_{Tra} \) and \( FF_{Dur} \) persist in individual language pairs?

• Does the global correlation between \( H_{Tra} \) and \( FF_{Dur} \) persist in individual translation modalities?

In other words, do macro-level findings about the correlation between \( H_{Tra} \) and \( FF_{Dur} \) still hold true at the micro-level?

For the present observational study, we chose to focus on a single translation modality featured in one TPR-DB study: the from-scratch translation task in the BML12 study (Mesa-Lao, 2014). The BML12 study also investigated and compared post-editing of machine-translated texts by translators and monolingual editing of machine-translated texts. The purpose of the present observational study on the BML12 data is to assess whether word translation entropy computed by restricting it to the from-scratch translation data continues to correlate with first fixation durations in the from-scratch translation task only. While this information complements the findings from Chapter 3 on the English-Spanish translation pair, there are essential differences: cross-linguistic ambiguity is measured differently; the translation direction is different; the translation task is different; and one study is experimental, but the other is correlational.

In this chapter, the observational study of word translation entropy in the BML12 dataset that is part of the CRITT TPR-DB is reported as follows: first, the CRITT TPR-DB is presented, then, previous studies conducted with the BML12 database subset are discussed, followed by the procedure, results and implications of the observational study conducted in the BML12 and last, the experimental data from Chapter 3 and the observation data in the present Chapter are compared and contrasted.
The CRITT TPR-DB

The Translation Process Research Database (TPR-DB) is an initiative emanating from the Center for Research and Innovation in Translation and Translation Technology (CRITT) at the Copenhagen Business School. This database is publicly available under a creative commons license and contains about thirty studies\(^8\) of translation, post-editing, revision, authoring and copying tasks during which eye-tracking and keystroke logging data were collected via Translog and the CASMACAT workbench\(^9\) (Carl et al., 2016b, p. 13). The overall data available amounts to over 500 hours of text production, in ten different target languages (ibid.). Within this database, the multiLing study aims at comparing from-scratch translation (T), post-editing (P) (comparison of source and target texts), monolingual post-editing (E) (editing of the target text only, without access to the source text), and copying (C) for different translators and language pairs (CRITT TPR-DB website). In this chapter, we will examine data on from-scratch translation in a single component, BML12, of the multiLing dataset of the TPR-DB. The BML12 data was collected from 20 translation sessions, 20 post-editing sessions and 20 editing sessions, completed by students and experienced translators, from English into Spanish. Six short English source texts were used and their lengths ranged from 110 to 160 words. Data was collected with a Tobii T120 (Carl, 2012).

\(^8\) Studies are presented on the CRITT TPR-DB website.
Studies Involving the BML12 Data Set

Mesa-Lao (2014) investigated gaze behavior in translation and post-editing in the English into Spanish direction. Six professional translators worked with four texts (compiled in the BML12 database) and measures of task time, number of fixations, total gaze time duration and transitions between the source and target areas were recorded. Results showed different behavioral patterns between translation and post-editing. More time and more fixations were recorded in translation, and half of the participants displayed more ST fixations while the other half displayed more TT fixations. Shorter time and overall more fixations on the target text were recorded in post-editing. The authors proposed two progression graphs to represent a prototypical translation task and a prototypical post-editing task. These progression graphs identify some essential differences between translation modalities. For example, in from-scratch translation it is typical for translators to briefly scan the source text, or a significant chunk of the source text depending on translator preference, before engaging in the production of the target text. During target text production, the eyes fixate alternately on localized relevant parts of both source and target texts as the translation unfolds, both for ST comprehension and TT production monitoring. Finally, the translator takes a global approach again during a revision phase that focuses more on the target text than on the source text. However, during post-editing, the initial scanning of the source text is largely missing because the post-editor primarily focuses on finding errors in the target text. Such differences are likely to influence patterns of first fixation durations in the source text, which may affect correlations between measures of cross-linguistic ambiguity and first fixation duration in different translation modalities.
We will use Mesa-Lao’s data to investigate the specific issue of whether conclusions drawn from the prior global analysis of the relationship between cross-linguistic ambiguity and first fixation duration to investigate early target language activation during reading for translation in multiple translation modalities and across several language pairs (Schaeffer et al., 2016b) still hold true in the context of the single modality of from-scratch translation from English to Spanish. However, it is instructive to outline a wider range of research whose data is encapsulated in the TPR-DB and that is relevant to the study of co-activation of languages during translation, variability in translation output, and cognitive effort expended in producing translation output.

Schaeffer et al. (2016b) argued that early eye-movements correspond to automatic processes whereas later ones correspond to conscious processing. Building on their perspective that the translation process is both vertical and horizontal (Schaeffer & Carl, 2013), they explained that ST understanding is facilitated by TT production (Danks & Griffin, 1997). The study differed from previous ones in that it focused on early eye-movements as indicators of automatic processes, showing the impact of target language specific items in ST reading. The data sets used in this study were ACS08, BD08, BD13, BML12, HLR13, KTHJ08, MS12, NJ12 and SG12, totaling 295 recordings.

The early eye movement measure of first fixation durations (FFDur) (first contact with a word) was significantly positively correlated with word translation entropy (a measure of TT variability at the word level) and cross-lingual distortion (a measure of the differences in word alignment between ST and TT). The authors argued that word translation entropy and cross-lingual distortion are indicators of how literal a translation is. The results further corroborated the literal translation hypothesis as the default mode of translating as evidenced by how initial TT
representations shared ST word order. Total reading time measures showed similar results to first fixation duration, indicating that shared representations, activated automatically early on in the reading process may be a reference used later to regenerate the ST in the target language for monitoring purposes, as a monolingual vertical process.

To investigate language co-activation and reading processes in translation, Bangalore et al. (2015) measured syntactic variation in translation and post-editing in five data sets, using source text monolingual copying as a control task. It was hypothesized that higher syntactic variation, another form of entropy, would imply higher cognitive effort, that is, the use of more cognitive resources (e.g.: working memory, namely a theoretical temporary store for processing information) are needed to accomplish a task, evidenced by longer reading times on the source text and longer typing times. Source texts in English were translated into Danish, German and Spanish. Verb valency (e.g.: transitivity/intransitivity), clause dependency (independent versus dependent) and voice (active or passive) were annotated in each clause. The data was analyzed along the dependent variables of total reading time and duration of coherent typing behavior, and the predictor variable of syntactic variation. Results indicated a difference between task modalities. They confirmed the hypothesis only for the translation task. In the post-editing task, the machine translation output may have primed the post-editor, cancelling the effect of syntactic variation on source text reading time observed in the translation task.

Schaeffer et al. (2016a) suggested using Activity Units as a new method to measure cognitive effort in translation. These units take into account the transfer activities between the source and the target text, such as monitoring the target text production while typing and gazing at the source text in parallel. Activity Units complement:
- The pause-word ratio (PWR) measure (Lacruz & Shreve, 2014), which predicts cognitive effort in translation from the ratio of the number of pauses to the number of words in the segment.
- The Translation Difficulty Index (TDI) measure (Mishra et al., 2013), which predicts the cognitive effort applied in translation based on the sum of ST and TT reading times.

This global study first showed that PWR and TDI correlate, using data from translation and post-editing segments, respectively from English into Chinese, Japanese, Danish, German and Spanish, and from English into Chinese, Japanese, German and Spanish, extracted from the KTHJ08, BML12, SG12, MS12, ENJA15 and NJ12 databases of the TPR-DB. To build on the discovery of this correlation and attempt to find a measure that would best represent the translation activity in terms of gaze behavior during translation typing pauses, the following types of Activity Units were proposed: 1) ST reading, 2) TT reading, 3) TT typing only, 4) ST reading and typing, 5) TT reading and typing and 6) no gaze or typing activity recorded in a 5-second time period. Shifts between activity types define Activity Units. The authors studied bigrams and trigrams of activity units and found that although post-editing required more iteration of ST-TT reading cycles compared with translation, gazing and typing behavior remained similar across both tasks. Specifically, it was shown that the longer and the less linear were gazes on the ST, the more likely were translators to subsequently read the produced TT. Similarly, the shorter and the more linear were ST gazes, the more likely were translators to start typing immediately after the ST reading unit.

Bangalore et al. (2016) studied how syntactic variation and priming vary in translation compared with the control task of monolingual copying, using the following behavioral measures as evidence: coherent typing activity per word, total ST and TT reading times per word and
average first fixation duration for each segment. The authors assumed a shared cognitive network across the two languages to explain how the involvement of the two languages in translation can lead to processing facilitations (priming effect). They use translation data from English into Danish, German and Spanish extracted respectively from KTHJ08, SG12 and BML12. Results showed that syntactic variation (which can be viewed as a form of entropy) predicted higher ST segment reading times and, to a lesser extent, higher average length of ST first fixation durations in translation compared with monolingual copying. In addition, syntactic variation correlated positively with typing inefficiency and total TT reading time. These results suggest that higher syntactic variation indicates higher cognitive effort, in turn supporting the view that languages are co-activated throughout the translation task, relying on shared lexical and syntactic representations across languages.

To delve deeper into translation literality and contribute to building understanding of why translation is difficult, Carl and Schaeffer (2017) reflected on the previous results on translation literality in Schaeffer and Carl (2013), Schaeffer et al. (2014; 2017) and Schaeffer et al. (2016b), and further examined syntactic and semantic distance between languages in translation and post-editing. The authors took an empirical approach using a multilingual subset of the TPR-DB of the same English STs translated and post-edited in Danish, German, Spanish, Hindi, Japanese and Chinese. They used the word translation entropy metric ($H_{Tra}$) that assesses the semantic distance between source and target text from the informational content of the translations of source text words. In addition, they used a metric, $H_{Cross}$, to gather information on word order choices made by translators to assess the syntactic distance between source and target texts. $H_{Tra}$ and $H_{Cross}$ are measures of translation literality and were assessed against the data in the TPR-DB in five languages (Danish, Spanish, German, Hindi and Japanese). Hindi and Japanese
both had higher $H_{Tra}$ and $H_{Cross}$ values compared with the other languages, which may be explained by the fact that they are more remote from English – the metrics provide empirical data to support this view. Moreover, in both translation and post-editing tasks, participants took longer and made more mistakes when typing or post-editing cross-linguistically highly ambiguous words. That is, words with higher word translation entropy values (higher information content in the translations proposed for that word) took longer to process than cross-linguistically less ambiguous words (words with lower entropy values). This effect was also observed in a post-editing task in which $H_{Tra}$ positively and significantly correlated with the time to produce a segment (Báez et al., 2017). However, in an indication of processing differences in different translation modalities, greater semantic variation was found in from-scratch translation tasks than in post-editing. This might be explained by post-editors being strongly influenced by the translation suggestions they see and therefore being more likely to accept them regardless of their quality. Such an influence and its potential consequences would not be present in from-scratch translation. Acceptance of suggested translations would thus reduce the level of translation variability in post-editing below the level in from-scratch translation.

Based on the $H_{Tra}$ and $H_{Cross}$ metrics and their effect on language processing, Schaeffer and Carl (2017) argued that the degree of literality of a translation can help predict when translators or post-editors will type following ST or TT reading. Their goal was to propose a minimal cognitive model for translating and post-editing based on empirical data gathered within the multilingual subset of the TPR-DB. The translation process was divided into four types of activities: ST reading, TT reading, writing with or without simultaneous ST or TT reading, and pausing (no activity for longer than 2.5 seconds). These four activities occur at
different times and recurrently both in the translation and post-editing process. By analyzing gaze and typing patterns, it was possible to predict activities in the translation and post-editing tasks. This revealed that the effect of word order entropy, measured by $H_{Cross}$, was similar in post-editing and from-scratch translation, pointing to common processes. In sum, the authors envision a cognitive model of translating and post-editing based on minimal activity units of problem solving and solution-finding.

**Observational Study of $H_{Tra}$ and $FFDur$ in BML12**

**Data Preparation**

Our observational study focused on the data collected from 32 participants carrying out from-scratch translation from English into Spanish of six source texts.

For the eye tracking studies in the CRITT database, the mean first fixation duration, $FFDur$, was computed for each distinct English word appearing in one or more of the source texts. In computing $FFDur$, data was taken from all participants engaged in all translation modalities from source language English into all available target languages. Punctuations, such as semi-colons and periods also had a $FFDur$ value.

In the present study, we only consider the from-scratch translations from English into Spanish recorded in BML12. Translation from English into Spanish may exhibit different source text reading characteristics from translation into other languages, such as Chinese, Danish, Dutch, German, Hindi, and Japanese, represented in the CRITT database. Such differences may, for example, be triggered by interactions between the source and target languages. In addition,
we have already noted evidence (Mesa-Lao, 2014) that reading progressions vary between different translation tasks. As a consequence, it is possible that first fixation durations have different characteristics related not just to differences between language pairs, but also related to different translation tasks. Accordingly, we computed first fixation duration means corresponding to FFDur but restricted to the from-scratch translation data in BML12. We refer to this from-scratch translation first fixation duration variable as TFFDur.

The other variable whose values we computed from the from-scratch translation data in BML12 is a variant of cross-linguistic entropy HTra introduced by Carl et al. (2016b). The notion of HTra was inspired by the concept of entropy H introduced by Shannon (1948) in his development of Information Theory; this, in turn, was related to the physical concept of entropy developed by Boltzmann (1872). In any of its manifestations, entropy is a measure of uncertainty about a situation where several unpredictable outcomes are possible. Each possible outcome carries information (I) about the situation that is related to the probability (p), or relative frequency, of the occurrence of that outcome. Through an argument beyond the scope of this dissertation (see Shannon 1948), the relationship between the information and relative frequency is mathematically constrained to satisfy \( I = \log(p) \). By convention, the logarithm is normally taken to the base 2. The information entropy \( H \) of a situation is the weighted average of the information associated with each outcome. This is computed by calculating the value of \( p*I \) for each outcome and adding up all such values. Entropy is zero when a single outcome is completely predictable: \( p = 1 \), so \( I = 0 \), and \( H = 0 \).

Word translation entropy is a measure of uncertainty about the translation of a source language word where several possible translations are possible into the target language. Each choice of translation carries information (I) about the translation situation that is related to the
probability \( (p) \), or relative frequency, of the occurrence of that choice in the way described above. After the relative frequencies of the observed translation choices for a source language word are determined, its translation entropy \( H_{Tra} \) can be computed analogously to the procedure described in the previous paragraph as the weighted average of the translation information associated with each translation choice. In the absence of translation errors, \( H_{Tra} = 0 \) for cross-linguistically non-ambiguous words.

Studies involving \( H_{Tra} \) based on the CRITT database generally aggregate word translation choices made in different translation modalities. However, as we noted in the discussion of \( TFFDur \), there are potentially large differences between different modalities in cognitive processing during source text reading. Since the modality we are considering in this Chapter is from-scratch translation, we work with a from-scratch variant \( TH_{Tra} \) of \( H_{Tra} \) obtained as the weighted average of the translation information associated with each from-scratch translation choice. Details of a typical computation of \( H_{Tra}/TH_{Tra} \) are given in Table 2.
### Table 2

**Calculation of THtra for the Source Language Word ‘imprisoned’ in BML12**

<table>
<thead>
<tr>
<th>Source text word: imprisoned</th>
<th>Translation instances: 11</th>
<th>Logarithms to base 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation version</td>
<td>Count (c)</td>
<td>Relative frequency $p = c/11$</td>
</tr>
<tr>
<td>encarcelado</td>
<td>4</td>
<td>0.364</td>
</tr>
<tr>
<td>condenado</td>
<td>2</td>
<td>0.182</td>
</tr>
<tr>
<td>condenado</td>
<td>1</td>
<td>0.090</td>
</tr>
<tr>
<td>prisión</td>
<td>4</td>
<td>0.364</td>
</tr>
</tbody>
</table>

$THtra$ (sum of contributions)  

$= 1.823$

### Correlational Study of BML12 Data

We carried out a detailed analysis within the BML12 subset of the CRITT database, in order to be able to focus on a single language pair and direction (English to Spanish) and to disaggregate the from-scratch translation data. Our analysis of the BML12 data is based on the metrics $HTra$ (word translation entropy) and $FFDur$ (first fixation duration). The analysis of the from-scratch translation subset of BML12 is based on the metrics $THTra$ (word translation entropy computed using the from-scratch translation task only) and $TFFDur$ (the mean first fixation duration, computed for each word across all participants in the from-scratch translation task). Since the distributions of $FFDur$ and $TFFDur$ are skewed to the right, we followed Schaeffer et al.’s (2016b) work with $FFDur$ in applying a logarithmic transformation to bring the distribution closer to a normal distribution and so more appropriate for a linear correlational
analysis. We label the logarithmically transformed versions of $FFDur$ and $TFFDur$ as $LogFFDur$ and $LogTFFDur$, respectively.

As part of their global analysis of the multiLing subset of the CRITT database, Schaeffer et al. (2016b) found a significant weak positive correlation between $Htra$ and $LogFFDur$. Thus, working with a different form of cross-linguistic ambiguity from the one we considered in Chapter 3, they found evidence for a positive association between cross-linguistic ambiguity and first fixation duration. Higher levels of cross-linguistic ambiguity ($Htra$) were associated with higher values of $LogFFDur$, and so longer first fixation durations while reading a source text as part of a translation task. This evidence of early activation of the target language during the process of a translation task suggests early horizontal processing.

Schaeffer et al.’s analysis involved several language pairs and multiple translation modalities. We now focus on a single direction in a single pair using the data from BML12, which is a record of various types of translation tasks from English into Spanish. Our first question was whether the association between cross-linguistic ambiguity and first fixation durations is still present for this single direction in this language pair. Measures on the editing task were not included in the computation because first fixation durations on the source text in this task were null for the overwhelming majority. This is explained by the fact that the monolingual editing task by definition does not or minimally includes access to the source text. Computation of the Pearson correlation coefficient for $Htra$ and $LogFFDur$ in BML12 demonstrated a significant, but extremely weak positive correlation ($r(844) = 0.069, p = 0.045$). In other words, as in a previous analysis by Schaeffer et al. (2016b) of the multiLing subset of the CRITT database, higher levels of cross-linguistic ambiguity ($Htra$) were associated with higher values of $LogFFDur$, meaning longer first fixation durations while reading an English
source text as part of a translation task into Spanish. The extreme weakness of the correlation provides only very weak evidence for early activation of Spanish during the process of a translation task from English into Spanish, and so only very weak evidence for early horizontal processing.

However, the weakness of the correlation could be explained by the fact that two BML12 tasks contribute to the \( H_{Tr} \) value, and both might not affect \( FFDur \) in the same way due to the potentially different cognitive requirements of each task as suggested in Mesa-Lao (2014), Bangalore et al. (2015) and Carl and Schaeffer (2017). The BML12 data includes from-scratch translation, post-editing of machine translation and monolingual post-editing of machine translation, but only the two first tasks were included in the correlation computation as explained above. In normal post-editing, the machine translation in the target language is edited with input from the source language, and so the production of the target text involves extra input as compared with the from-scratch translation process. We next attempted to disentangle the effects of the different translation modalities by focusing on the single task of from-scratch translation.

As from-scratch translation is the task used in Chapter 3, this extra focus allows more direct comparisons with the results from Chapter 3. We recomputed \( H_{Tr} \) taking into account only the English translations that appeared in the from-scratch translations; we labeled this variable as \( T_{HTr} \). Then we computed the means of the logarithms of the first fixation durations for each source text word, and we labeled this variable \( LogTFFDur \). Computation of the Pearson correlation coefficient for \( T_{HTr} \) and \( LogTFFDur \) in BML12 showed no significant correlation between these variables \((r(844) = 0.023, p = .507)\).

In view of the results of the Experiment in Chapter 3, this negative result is not altogether surprising. It does lend further support to the possibility that different translation modalities
require different early reading processing. We will consider this possibility further in the Discussion below. However, in order to be able to make more direct comparisons of the experimental and observational approaches, we first re-examine the data from the Experiment from the perspective of word translation entropy.

**Correlational Study of Experimental Data**

We assessed whether the materials we used in Chapter 3 would yield similar findings about first fixation durations and word translation entropy as those stemming from the BML12 data. To achieve this, $THTra$ measures and $LogTFFDur$ averages were computed from the data obtained in Experiment 3. It should be noted that although the language pair is the same in both situations, the translation directions are opposite. In addition, the experimental task was sight translation, whereas the task in the BML12 was from-scratch translation, involving slightly different processes as discussed previously.

We began by determining whether the work of the translators in the Experiment resulted in a significant difference in $THTra$ values for the cross-linguistically ambiguous and cross-linguistically non-ambiguous stimuli. First Fixation Duration was greater in the Ambiguous condition ($M = 1.191$, $SD = 0.814$) than in the Non-Ambiguous condition ($M = 0.490$, $SD = 0.110$); $t(19) = 4.050$, $p = 0.001$. However, as in BML12, there was no significant correlation between $THTra$ and $LogTFFDur$ for the experimental materials ($r(38) = 0.143$, $p = 0.379$).
Discussion

Overall, on the micro-scale level of Mesa-Lao’s (2014) study of translation tasks from English into Spanish, a weak, but significant correlation was found between $H_{Tra}$ and the log transformation of $FF_{Dur}$, paralleling the macro-scale results found in Schaeffer et al. (2016b). However, the from-scratch-translation-only measure of translation word entropy, $T_{HTra}$, did not significantly correlate with the log transformation of the from-scratch translation-only version, $T_{FF_{Dur}}$, of $FF_{Dur}$. This suggests that the effect found at the macro-scale is not present at this particular micro-scale. Such a difference between the from-scratch translation task in isolation and the aggregated translation tasks in BML12 (post-editing of machine-translated text, monolingual editing and from-scratch translation) suggests that different tasks trigger different early reading processes. Although less is known about early processing differences, overall processing differences between the tasks are well established: post-editing of machine-translated text with access to the source is less effortful, but requires more switches between source and target texts, leading to overall shorter fixations on the target text, while from-scratch translation is more effortful, requires more overall time and more fixations (Mesa Lao, 2014). One possible reason why early processing differences might arise is that post-editors, who are expected to find and correct machine translation errors, are predisposed to expect ambiguity in the form of higher entropy due to errors. Such a mindset might be expected to exert an influence on early processing in post-editing that would be absent in from-scratch translation.

There are also good reasons to expect differences in the translation word entropy that results in different translation tasks. Bangalore et al. (2015) argued that the effect of entropy is diminished in post-editing compared with from-scratch translation because the solutions
provided by the machine prime the post-editor. This suggests that there is less solution-finding activity, or even creativity in post-editing, and thus less translation variation, because from the very beginning the task leads the post-editor to assess an already-generated solution, as opposed to generating one from scratch. This argument is also supported by the findings in Báez et al. (2017) and Carl and Schaeffer (2017): Báez et al. do find a significant positive correlation between translation word entropy and overall time spent on a segment in post-editing, but according to Carl and Schaeffer (2017), the entropy in post-editing is lower than in from-scratch translation, which could again be partially attributed to early priming effects of machine translation solutions on the translator.

A further interpretation of the results we found might be based on previous findings that the Cross measure (Schaeffer et al., 2016b) of syntactic variation between source and target texts plays a significant role in early reading processes, as evidenced by first fixation durations. Indeed, Schaeffer et al. found that Cross, which can be viewed as a form of syntactic entropy in translation, is significantly positively correlated with LogFFDur. Further research on the impact of Cross in analyzing the experimental data and the translation subset of the BML12 would be needed to assess the validity of this possible interpretation.

We have outlined preliminary evidence suggesting differences between early reading processes during from-scratch translation and post-editing. If such differences are confirmed, it may be necessary to re-evaluate previous macro-scale research to take into account the implications of task differences, consistent with the models proposed by Schaeffer and Carl (2017), and to conduct further research to better understand early reading processes in different translation modalities. As evidence accumulates, general tendencies will be able to emerge, which will strengthen the foundations of translation process research.
CHAPTER 5

CONCLUSIONS

Summary of Main Results

This study first aimed to experimentally assess the time-course of early dual-language activation in translation by analyzing first fixation durations on cross-linguistically ambiguous words versus cross-linguistically non-ambiguous words in a sight translation task from Spanish into English. The study was then complemented by a correlational investigation of a broader definition of cross-linguistic ambiguity, including word translation entropy as a marker of semantic ambiguity in translation.

Experimental results showed that, contrary to initial expectations and previous results in the monolingual and bilingual literature on ambiguity resolution, first fixation duration were not significantly different on cross-linguistically ambiguous words than on cross-linguistically non-ambiguous words. This result was attributed to the high constraints imposed on the selection of stimuli for the Experiment. In addition, participants’ language combination (either L1 Spanish and L2 English or vice-versa) did not significantly affect first fixation durations.

To gain insight into the early processes involved in reading for translation, a new approach was taken using the CRITT database, focusing on the analysis of first fixation durations on cross-linguistically ambiguous words in a from-scratch translation task from English into Spanish, in the
BML12 subset of the database. Results of this micro-scale study showed a weak significant correlation between word translation entropy and the log transformation of first fixation durations across translation modalities (from-scratch translation, post-editing of machine translation output and monolingual editing of post-edited machine translation output), in line with previous macro-scale findings.

When the word translation entropy measure was isolated to take into account the from-scratch translation modality only, no significant correlation was found with the log transformation of first fixation durations. The macro-scale results were therefore not replicated at this micro-scale. This result was attributed to processing differences across translation modalities. Specifically, post-editors’ early stage reading behavior might be affected by the expectation of finding ambiguity, or higher entropy, due to errors in the machine translation output. This mindset is absent in from-scratch translation. In addition, because machine translation output primes the post-editor to assess a solution as opposed to generating one, post-editing would manifest less translation variation than from-scratch translation.

**Implications**

**Modeling Reading for Translation**

Experimental results and results from the analysis of BML12 suggest that modeling reading for translation can benefit from at least a two-pronged approach: experimental and correlational. Moreover, these results show the necessity to disentangle the different processes and effects resulting from various translation modalities. This work also contributed to modeling reading for
translation by showing a weak significant correlation at the micro-scale between word translation entropy and the log transformation of first fixation durations across translation modalities, replicating previous findings at the macro-scale. Previous studies (Schaeffer et al., 2014; 2017) suggest that the two languages involved in the translation task are co-activated (horizontal processing), with a monitoring process occurring only later (vertical process) to assess the quality of the produced translation. The time-course of language co-activation is further refined in Schaeffer et al. (2016b), who used the word translation entropy measure to find that language co-activation takes place early in reading for translation. Horizontal and vertical processes seem to manifest themselves in alternation as suggested by the recursive model of reading for translation (Schaeffer & Carl, 2013).

**Pedagogy**

Because the translator must formulate a text in a target language from an already existing text in the source language, the reading of the source text must be done actively, that is with close attention to both the text’s form and content and one’s knowledge about it. Active reading allows for the process to be efficient and effective: the interpretation of the text is accurate and arrived at in a minimal amount of time. However, strategies for effective reading for translation (Washbourne, 2012) are not taught in translator training. Indeed, reading in L1 or L2 is taken as an acquired skill, and therefore not taught explicitly. Reading for translation however is a developing skill at this stage and students would benefit from specific tasks to hone their competences (e.g.: let students compare and contrast how they read for different purposes, discuss the relevant
background knowledge necessary to understand a given text for translation, find areas of interest in the source text to focus the attention during the translation process) (Neveu, in revision).

This study defined translation ambiguity controlling for two components: source language ambiguity resulting from polysemy and word translation entropy. The analysis of a subset of the CRITT database showed a weak significant correlation between word translation entropy and the log transformation of first fixation durations in different translation modalities at a micro-level, complementing previous macro-scale results that showed early co-activation of languages in translation. Students could potentially benefit from explicit knowledge of the potential effects of language co-activation, and, through instruction, they could learn to read more effectively for translation by dedicating more attention to parts of the text that could be more difficult as a result of entropy or source language polysemy. Enhanced awareness of the translation process in general has been shown to positively impact students’ translation processing and, as a result, to improve the quality of their translation products (Gross, 2003; Kiraly, 2000; Kussmaul, 1995).

Further Research

To complement micro-scale findings in the BML12 data set, the impact of the Cross measure, a form of syntactic entropy, could be studied in further investigations of early processes in reading for translation. Indeed, Cross has been found to be significantly positively correlated with the log transformation of first fixation duration in Schaeffer et al. (2016b).

Moreover, the different results found for early reading processes in from-scratch translation and post-editing suggest the need to take task modality differences into account in further macro-scale studies involving various subsets of the CRITT database. This would in turn contribute to
recent efforts aiming at modeling from-scratch translation and post-editing processes (Schaeffer & Carl, 2017).

As in the current investigation, controlled experimental research in Translation Process Research is hampered by the short supply of appropriate experimental stimuli and participants. Nevertheless, to eventually provide solid foundations for a robust model of translation processes, future experimental studies of early processes in reading for translation are needed to replicate and generalize recent findings. For example, it would be interesting to undertake a careful comparison of source language polysemy and entropy to evaluate the extent to which they have similar or different effects at different stages of reading for translation. Moreover, to better understand the reading process in translation, it would be relevant to include in experimental studies languages with different syntactical orders and scripts. In this context, it is important to note that Carl and Schaeffer (2017) have shown that lexical and syntactic entropy measures are affected differently by European and Asian languages. Further experimental research could also explore the effect of directionality on the translation of entropic words, with and without constraining context to assess the potential effects of language proficiency and context on processing ambiguity in reading for translation.


101


APPENDIX A

Table A1

*Ambiguous Spanish Words and their two Translations, with Frequency Ranking, Length and Mean Values*

<table>
<thead>
<tr>
<th>Ambiguous Spanish word</th>
<th>Frequency</th>
<th>Length</th>
<th>English translation #1</th>
<th>Frequency</th>
<th>Length</th>
<th>English translation #2</th>
<th>Frequency</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>abrigo</td>
<td>2996</td>
<td>6</td>
<td>coat</td>
<td>2284</td>
<td>4</td>
<td>shelter</td>
<td>2914</td>
<td>7</td>
</tr>
<tr>
<td>banco</td>
<td>881</td>
<td>5</td>
<td>bank</td>
<td>701</td>
<td>4</td>
<td>bench</td>
<td>2808</td>
<td>5</td>
</tr>
<tr>
<td>bolsa</td>
<td>1915</td>
<td>5</td>
<td>bag</td>
<td>1045</td>
<td>3</td>
<td>stock exchange&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1521</td>
<td>13</td>
</tr>
<tr>
<td>cañon</td>
<td>3229</td>
<td>5</td>
<td>canyon</td>
<td>5000&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6</td>
<td>cannon</td>
<td>5000</td>
<td>6</td>
</tr>
<tr>
<td>cita</td>
<td>2045</td>
<td>4</td>
<td>appointment</td>
<td>3084</td>
<td>11</td>
<td>quotation</td>
<td>5000</td>
<td>9</td>
</tr>
<tr>
<td>cola</td>
<td>1589</td>
<td>4</td>
<td>line</td>
<td>283</td>
<td>4</td>
<td>tail</td>
<td>3002</td>
<td>4</td>
</tr>
<tr>
<td>dedo</td>
<td>1248</td>
<td>4</td>
<td>finger</td>
<td>1067</td>
<td>6</td>
<td>toe</td>
<td>5000</td>
<td>3</td>
</tr>
<tr>
<td>ensayo</td>
<td>1835</td>
<td>6</td>
<td>essay</td>
<td>2585</td>
<td>5</td>
<td>rehearsal</td>
<td>5000</td>
<td>9</td>
</tr>
<tr>
<td>espina</td>
<td>3708</td>
<td>6</td>
<td>bone</td>
<td>1487</td>
<td>4</td>
<td>thorn</td>
<td>5000</td>
<td>5</td>
</tr>
<tr>
<td>estación</td>
<td>1275</td>
<td>8</td>
<td>season</td>
<td>535</td>
<td>6</td>
<td>station</td>
<td>841</td>
<td>7</td>
</tr>
<tr>
<td>ficha</td>
<td>3736</td>
<td>5</td>
<td>index card</td>
<td>5000</td>
<td>9</td>
<td>token</td>
<td>5000</td>
<td>5</td>
</tr>
<tr>
<td>firma</td>
<td>1567</td>
<td>5</td>
<td>company</td>
<td>195</td>
<td>7</td>
<td>signature</td>
<td>3857</td>
<td>9</td>
</tr>
<tr>
<td>hoja</td>
<td>1000</td>
<td>4</td>
<td>leaf</td>
<td>1613</td>
<td>4</td>
<td>sheet</td>
<td>1815</td>
<td>5</td>
</tr>
<tr>
<td>muñeca</td>
<td>3082</td>
<td>6</td>
<td>wrist</td>
<td>3696</td>
<td>5</td>
<td>doll</td>
<td>3820</td>
<td>4</td>
</tr>
<tr>
<td>papel</td>
<td>277</td>
<td>5</td>
<td>role</td>
<td>460</td>
<td>4</td>
<td>paper</td>
<td>533</td>
<td>5</td>
</tr>
<tr>
<td>pata</td>
<td>2052</td>
<td>4</td>
<td>leg</td>
<td>880</td>
<td>3</td>
<td>paw</td>
<td>5000</td>
<td>3</td>
</tr>
<tr>
<td>pico</td>
<td>1553</td>
<td>4</td>
<td>peak</td>
<td>2431</td>
<td>4</td>
<td>beak</td>
<td>5000</td>
<td>4</td>
</tr>
<tr>
<td>pista</td>
<td>2065</td>
<td>5</td>
<td>clue</td>
<td>3401</td>
<td>4</td>
<td>trace</td>
<td>4239</td>
<td>5</td>
</tr>
<tr>
<td>receta</td>
<td>3196</td>
<td>6</td>
<td>recipe</td>
<td>2626</td>
<td>6</td>
<td>prescription</td>
<td>3427</td>
<td>12</td>
</tr>
<tr>
<td>vela</td>
<td>2287</td>
<td>4</td>
<td>candle</td>
<td>3948</td>
<td>6</td>
<td>sail</td>
<td>5000</td>
<td>4</td>
</tr>
</tbody>
</table>

| Mean                    | 2076.80   | 5.05   | 2116.05                 | 5.25      | 3688.85 | 6.20                    |

<sup>Note</sup>. Words with higher frequency rankings will tend to have lower frequencies in language.

<sup>a</sup>For frequencies > 5000, the COCA was used to determine the most frequent translation. Frequencies greater than 5000 are recorded as 5000. <sup>b</sup>Frequency of stock exchange recorded as that of exchange (more frequent than stock).
Table A2

Non-Ambiguous Spanish Words and their Translation, with Frequency Ranking, Length and Mean Values

<table>
<thead>
<tr>
<th>Matched non-ambiguous Spanish word</th>
<th>Frequency</th>
<th>Length</th>
<th>English translation</th>
<th>Frequency</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>mente</td>
<td>993</td>
<td>5</td>
<td>mind</td>
<td>477</td>
<td>4</td>
</tr>
<tr>
<td>negocio</td>
<td>864</td>
<td>7</td>
<td>business</td>
<td>254</td>
<td>8</td>
</tr>
<tr>
<td>pájaro</td>
<td>1824</td>
<td>6</td>
<td>bird</td>
<td>1176</td>
<td>4</td>
</tr>
<tr>
<td>pollo</td>
<td>3231</td>
<td>5</td>
<td>chicken</td>
<td>1596</td>
<td>7</td>
</tr>
<tr>
<td>mito</td>
<td>2467</td>
<td>4</td>
<td>myth</td>
<td>3087</td>
<td>4</td>
</tr>
<tr>
<td>pelo</td>
<td>1056</td>
<td>4</td>
<td>hair</td>
<td>618</td>
<td>4</td>
</tr>
<tr>
<td>brazo</td>
<td>620</td>
<td>5</td>
<td>arm</td>
<td>495</td>
<td>3</td>
</tr>
<tr>
<td>canción</td>
<td>1459</td>
<td>7</td>
<td>song</td>
<td>908</td>
<td>4</td>
</tr>
<tr>
<td>calle</td>
<td>248</td>
<td>5</td>
<td>street</td>
<td>352</td>
<td>6</td>
</tr>
<tr>
<td>montaña</td>
<td>1212</td>
<td>7</td>
<td>mountain</td>
<td>852</td>
<td>8</td>
</tr>
<tr>
<td>guante</td>
<td>3392</td>
<td>6</td>
<td>glove</td>
<td>3080</td>
<td>5</td>
</tr>
<tr>
<td>regreso</td>
<td>1557</td>
<td>7</td>
<td>return</td>
<td>1274</td>
<td>6</td>
</tr>
<tr>
<td>camisa</td>
<td>2443</td>
<td>6</td>
<td>t-shirt</td>
<td>3594</td>
<td>6</td>
</tr>
<tr>
<td>tinta</td>
<td>3089</td>
<td>5</td>
<td>ink</td>
<td>5000*</td>
<td>3</td>
</tr>
<tr>
<td>dinero</td>
<td>291</td>
<td>6</td>
<td>money</td>
<td>236</td>
<td>5</td>
</tr>
<tr>
<td>libro</td>
<td>253</td>
<td>5</td>
<td>book</td>
<td>245</td>
<td>4</td>
</tr>
<tr>
<td>huevo</td>
<td>1990</td>
<td>5</td>
<td>egg</td>
<td>1445</td>
<td>3</td>
</tr>
<tr>
<td>ropa</td>
<td>1285</td>
<td>4</td>
<td>clothes</td>
<td>1458</td>
<td>7</td>
</tr>
<tr>
<td>falda</td>
<td>3143</td>
<td>5</td>
<td>skirt</td>
<td>3754</td>
<td>5</td>
</tr>
<tr>
<td>cajón</td>
<td>2620</td>
<td>5</td>
<td>drawer</td>
<td>4244</td>
<td>6</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>1701.85</strong></td>
<td><strong>5.45</strong></td>
<td></td>
<td><strong>1707.25</strong></td>
<td><strong>5.10</strong></td>
</tr>
</tbody>
</table>

*Note. Words with higher frequency rankings will tend to have lower frequencies in language.

*aFor frequencies > 5000, the COCA was used to determine the most frequent translation. Frequencies greater than 5000 are recorded as 5000.
APPENDIX B

Experimental Sentences with Ambiguous (A) and Non-Ambiguous (C) Words

1. (A) En el lado derecho tenía un dedo roto cuando regresó a casa después del viaje.
2. (C) En el lado derecho tenía un brazo roto cuando regresó a casa después del viaje.

3. (A) A la chica le gustó mucho la bolsa que había visto en la televisión.
4. (C) A la chica le gustó mucho el pájaro que había visto en la televisión.

5. (A) No puedo encontrar por ninguna parte en la casa la receta que le hace falta a mi abuela para esta tarde.
6. (C) No puedo encontrar por ninguna parte en la casa la falda que le hace falta a mi abuela para esta tarde.

7. (A) Hay que tener mucho cuidado con las espinas para no acabar en el hospital.
8. (C) Hay que tener mucho cuidado con las calles para no acabar en el hospital.

9. (A) La secretaria me explicó que el escritor decidió utilizar el papel de la agencia en esta ocasión.
10. (C) La secretaria me explicó que el escritor decidió utilizar el dinero de la agencia en esta ocasión.

11. (A) Me gustan las grandes ciudades porque siempre hay muchos bancos que son muy cómodos.
12. (C) Me gustan las grandes ciudades porque siempre hay muchos negocios que son muy cómodos.

13. (A) Al llegar con mi familia, podíamos ver la larga cola y todo el mundo parecía sorprendido.
14. (C) Al llegar con mi familia, podíamos ver el largo pelo y todo el mundo parecía sorprendido.

15. (A) En esta ciudad, la estación es fría, así que hay que prepararse de antemano.
16. (C) En esta ciudad, la montaña es fría, así que hay que prepararse de antemano.

17. (A) No se podía recoger la hoja del suelo porque el viento soplabo con mucha intensidad.
18. (C) No se podía recoger la camisa del suelo porque el viento soplabo con mucha intensidad.

19. (A) Hoy compré varias cosas e incluso una nueva vela que podré utilizar más de una vez.
20. (C) Hoy compré varias cosas e incluso un nuevo cajón que podré utilizar más de una vez.

21. (A) María se sentía muy mal porque rompió la muñeca en la tienda y nadie se dio cuenta.
22. (C) María se sentía muy mal porque rompió la tinta en la tienda y nadie se dio cuenta.
23. (A) Aún desde aquí, se puede ver que el pico es muy bonito e impresionante.
24. (C) Aún desde aquí, se puede ver que el huevo es muy bonito e impresionante.

25. (A) Durante la reunión que tuvo lugar ayer, pidieron la firma de esas tres personas, lo que se hizo muy rápidamente.
26. (C) Durante la reunión que tuvo lugar ayer, pidieron el regreso de esas tres personas, lo que se hizo muy rápidamente.

27. (A) En la entrevista el director dijo que hablarán del ensayo en los primeros días de enero.
28. (C) En la entrevista el director dijo que hablarán de la canción en los primeros días de enero.

29. (A) Tenemos que hablar con el club de lectura, porque una cita no es suficiente para mejorar la calidad de nuestra presentación.
30. (C) Tenemos que hablar con el club de lectura, porque un mito no es suficiente para mejorar la calidad de nuestra presentación.

31. (A) Tuve que llamar a mi amigo para que arreglara la pata que estaba rota desde ayer por la noche.
32. (C) Tuve que llamar a mi amigo para que arreglara el libro que estaba roto desde ayer por la noche.

33. (A) Los dos hermanos fueron al parque y siguieron la pista para encontrar a la persona que había desaparecido por la mañana.
34. (C) Los dos hermanos fueron al parque y siguieron la ropa para encontrar a la persona que había desaparecido por la mañana.

35. (A) Me he dado cuenta que en caso de tormenta, ese nuevo abrigo es muy útil al hablar con mis amigos.
36. (C) Me he dado cuenta que en caso de tormenta, esa nueva mente es muy útil al hablar con mis amigos.

37. (A) La presentación claramente demostró que estar cerca de un cañón siempre supone un riesgo para los turistas.
38. (C) La presentación claramente demostró que estar cerca de un pollo siempre supone un riesgo para los turistas.

39. (A) Pasé todo el día en la oficina buscando la ficha pero creo que la he perdido.
40. (C) Pasé todo el día en la oficina buscando el guante pero creo que lo he perdido.

**Filler Sentences**

1. Mis amigos han visitado muchas ciudades en el mundo donde han conocido gente muy interesante.
2. Lo que me gusta más en este país son los festivales que ocurren regularmente durante el año.
3. El autor explicó que eligió esta palabra en el título de su nueva novela porque significa 'luz'.
4. Este fin de semana voy de senderismo con un amigo mío para relajarnos antes de los exámenes.
5. Para hacer la reserva, me hace falta el número del hotel y nuestro programa de vacaciones.
6. Intenté hablar con la secretaria pero dejé un mensaje ya que estaba fuera de la oficina a esta hora.
7. Hacemos este ejercicio en clase para medir el nivel de lógica de los estudiantes antes de empezar el curso.
8. La meta de esa organización es cuidar de los caballos viejos que ya no pueden participar a competiciones.
9. Esta es una situación personal delicada y hay que mantener una actitud serena y positiva hasta que se resuelva.
10. Los presidentes de los dos países están de acuerdo para firmar un acuerdo de paz y poner fin a tres años de guerra.
11. Hace veinte años cuando vivíamos en esta ciudad, este árbol ya estaba aquí pero mucho más pequeño.
12. Para llegar a la granja, hay que seguir la carretera hacia el norte y girar a la izquierda en el semáforo.
13. El pueblo está situado cerca de las montañas, en un valle dónde hay un río que provee la mayoría del agua.
14. Está película recrea muy bien el pasaje en el libro de la batalla principal entre los dos ejércitos.
15. Aunque la temperatura de hoy es muy alta, hace frío en la sala a causa del aire acondicionado.
16. Durante el verano, se puede ver esa estrella muy brillante que forma parte de la constelación de Orión.
17. Vamos a preparar una fiesta para el cumpleaños de Pablo con una sorpresa que espero le gustarás.
18. El parque del centro de la ciudad tiene un puente japonés muy bonito cerca del que me gusta sentarme a leer.
19. Aunque nunca tuvo una mascota, mi tía decidió comprar un gato para tener compañía.
20. Desde la torre del castillo, tenemos una vista increíble sobre el paisaje que rodea esta ciudad costera.
APPENDIX C

Demographic and Language History Questionnaire Results

A summary of the results from the Demographic History Questionnaire (DHQ) and the Language History Questionnaire (LHQ) is presented in Table C1.

Table C1

*Linguistic Experiences and Self-Assessed Proficiency Ratings of the Translation Student Participants (n=12) Working with English and Spanish*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean translation experience (months)</td>
<td>5.2</td>
</tr>
<tr>
<td>Mean age started learning L2</td>
<td>11</td>
</tr>
<tr>
<td>Mean number of years practicing L2</td>
<td>15</td>
</tr>
<tr>
<td>Mean L2 immersion experience (months)</td>
<td>31</td>
</tr>
<tr>
<td><em>Self-assessed ratings in L2 (English or Spanish)</em></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>5.8</td>
</tr>
<tr>
<td>Speaking</td>
<td>5.8</td>
</tr>
<tr>
<td>Reading</td>
<td>6.1</td>
</tr>
<tr>
<td>Writing</td>
<td>5.7</td>
</tr>
</tbody>
</table>

*a One outlier was removed from this mean. b Likert scale, 1 = very poor, 7 = native-like*
Demographic History Questionnaire Results

The DHQ allowed the acquisition of general information and translation-specific information tailored to the pool of participants at Kent State University in the Spanish track of the M.A. in Translation or in the Translation Studies Ph.D.

The participants’ ages ranged from 21 to 35, with half the participants in the 21-25 year-old category. Two participants were doctoral students, nine were master’s students in their first or second year and one was a graduate of the Master’s program and was currently a Spanish instructor.

All participants with L1 English had United States (U.S.) nationality except for one participant who had dual U.S. and United Kingdom nationality. Four participants with L1 Spanish had South American nationalities, one had Spanish nationality, and one had dual U.S.-Spanish nationality.

All participants obtained a degree in at least one field related to languages prior to entering the M.A. in Translation at Kent State University. Five had no prior translation experience. Of the seven participants with prior translation experience, two had been freelancing and had completed internships, three had been freelancing only, one had interned only and one had work experience only.

A third of the participants reported translating from and into both their L1 and L2 at home, for themselves, at school and at work. A quarter only translated into one of their languages in one of the situations. One participant reported not translating in any of the situations.
Language History Questionnaire Results

The Language History Questionnaire (LHQ) served the purpose of getting participants’ linguistic background, to assess the homogeneity of the sample.

**Duration of L2 exposure.** All but three participants began learning their L2 after age 10. This means that most participants learned their L2 around the still debated critical period, which may mean that their proficiency is not as high as it would have been if the L2 had been acquired in early childhood (DeKeyser & Larson-Hall, 2005). A quarter of the participants reported fewer than 10 years of L2 study, so, as they were training to become professional translators, most participants had more than ten years L2 practice. It is interesting to point out that according to Ericsson, Krampe and Tesch-Römer (1993), expertise is attained after 10 years or 10,000 hours of deliberate practice (see Shreve, 2006 for applications in translation): this suggests that most participants had developed some expertise in translation when they did the experiment, supporting the homogeneity of the sample.

**Intensity of L2 exposure.** While two participants had extensive immersion (averaging ten years) in a country where their L2 was spoken, seven had a much shorter immersion experience (less than ten months), and three had experience around the mean of thirty months. No participant had zero immersion experience, which can be explained by the fact that students preparing to be professional translators would likely have mastered a second language by traveling to a country where it is spoken.
**Self-assessment of L2 proficiency.** In line with participants’ L2 experience and immersion, all self-rated at least 5 (*good*) on listening comprehension, with the exception of one participant at 4 (*functional*); the average self-rated proficiency level was 5.8 (*very good*).

As with listening comprehension, participants self-rated high on the speaking scale; listening and speaking skills are particularly enhanced while immersed in the L2 as the ear becomes acclimated to the accent and the subject needs to speak the language to be understood. The mean speaking score was 5.8 (*very good*).

All participants self-rated 5 or higher on the writing ability scale, with the exception of one participant who self-rated at the level of 3 (limited). The mean self-assessed writing ability rating was 5.7. While writing ability is highly important in translation, which explains the overall high self-rating score on this measure, this ability continues to be developed during training, which explains the lower self-rating score for one participant.

All participants self-rated 5 or higher on the reading ability scale. The mean for all participants was 6.1. Reading and writing ability go together in translation, therefore these scores are consistent with the writing scale scores, with the exception of one participant who self-rated more proficient in reading than writing in L2.
APPENDIX D

Table D1

*Data Collected per Participant as a Mean of all First Fixation Durations (FFDur) on Ambiguous and Non-Ambiguous Words in the Sight Translation Task in Milliseconds, Depending on the Native Language (L1)*

<table>
<thead>
<tr>
<th>Participant</th>
<th>L1</th>
<th>Mean FFDur Ambiguous</th>
<th>Mean FFDur Non-Ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>P03A</td>
<td>SP</td>
<td>167</td>
<td>147</td>
</tr>
<tr>
<td>P05A</td>
<td>SP</td>
<td>287</td>
<td>80</td>
</tr>
<tr>
<td>P15B</td>
<td>SP</td>
<td>333</td>
<td>747</td>
</tr>
<tr>
<td>P16B</td>
<td>SP</td>
<td>150</td>
<td>217</td>
</tr>
<tr>
<td>P10A</td>
<td>SP</td>
<td>120</td>
<td>197</td>
</tr>
<tr>
<td>P06B</td>
<td>SP</td>
<td>7</td>
<td>327</td>
</tr>
<tr>
<td>P07A</td>
<td>ENG</td>
<td>169</td>
<td>181</td>
</tr>
<tr>
<td>P09A</td>
<td>ENG</td>
<td>390</td>
<td>630</td>
</tr>
<tr>
<td>P14A</td>
<td>ENG</td>
<td>167</td>
<td>177</td>
</tr>
<tr>
<td>P08B</td>
<td>ENG</td>
<td>223</td>
<td>263</td>
</tr>
<tr>
<td>P13B</td>
<td>ENG</td>
<td>620</td>
<td>243</td>
</tr>
<tr>
<td><strong>Overall means</strong></td>
<td></td>
<td><strong>239</strong></td>
<td><strong>292</strong></td>
</tr>
</tbody>
</table>