Computer-Assisted Translation: An Empirical Investigation of Cognitive Effort

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

by

Christopher Davey Mellinger

May, 2014

Dissertation written by
Christopher D. Mellinger
B.A., Wake Forest University, 2007
M.A., Kent State University, 2009
Ph.D., Kent State University, 2014

Approved by

Keiran Dunne, Chair, Doctoral Dissertation Committee

Gregory Shreve, Members, Doctoral Dissertation Committee

Erik Angelone

Jonathan Maletic

William Merriman

Accepted by

Keiran Dunne, Chair, Department of Modern and Classical Language Studies

Raymond Craig, Dean, College of Arts and Sciences

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express sincere thanks to my advisor, Dr. Keiran Dunne, for his support and insightful comments throughout the entire dissertation process. His encouragement and experience were most valuable during my time in the doctoral program. I would also like to thank my dissertation committee members, Dr. Gregory Shreve, Dr. Erik Angelone, Dr. Jonathan Maletic, and Dr. William Merriman for their thought-provoking questions and suggestions during the course of my work. Many thanks to the faculty of the Department of Modern and Classical Language Studies and the Institute for Applied Linguistics for challenging me to grow as a teacher, researcher, and student. I am deeply grateful for the love and encouragement of my parents, brothers, and family. Their unwavering support is much appreciated. Thank you to all of my fellow classmates and friends who have shared this experience along the way. Lastly, many thanks to the professional translators who participated in this study.

Christopher Davey Mellinger

Kent, Ohio

May 9, 2014

CHAPTER 1

Introduction

Global commerce is an increasingly complex endeavor and requires companies to adapt traditional business strategies used domestically in order to reach diverse international markets. More than ever, the ways in which businesses communicate internally and with potential or actual clients must now incorporate strategies for multilingual and multicultural communication. In addition, companies must address exponential increases in the amount of digital content that is created. An IDC 2008 white paper indicated that 281 exabytes of data existed in the digital universe, and a more recent infographic (EMC 2011) showed that "1.2 zettabytes of data was created and replicated in 2010." This content—audio, video, photo, or text, or any other material that can be represented on a computer—is constantly being created, managed, stored, and transformed. Managing this content is a daunting undertaking, and requires immense resources; in fact, "since 2005, the investment by enterprises in the digital universe has increased 50% to $4 trillion – money spent to create, manage, store and derive revenues from the digital universe" (EMC 2011). Companies have placed a sizeable emphasis on generating revenues from this digital content, and yet increases in the number of IT professionals managing this content to the actual amount of data being managed is outpaced almost fifty-fold.

Written content is one communication channel that companies use internally and with clients. Given the multilingual nature of international commerce, written content must often appear in multiple languages.[1] Most companies, though, do not specialize in the transformation of this content from one language into another, nor are there in-house units dedicated to its creation. A dearth of experience and expertise in this area has, in part, given rise to the language services industry, which in 2012 was approximately valued at $35.5 billion dollars (Kelly, DePalma, and Stewart 2012:3). Language service providers (LSPs) often are the outsourced subcontractors specializing in the language services required by companies. Translation is one of the most-offered services, and specifically addresses the transformation of written content in specific languages. Nevertheless, the staggering difference between the relatively modest financial resources allocated to content management and the total size of the language service industry indicates that only a fraction of the written material is ever translated. Even so, the exponential growth of content shows a constant and growing need for language services.

To meet the ever-increasing volume of digital content requiring translation, LSPs and software developers have implemented various technologies into their work environments. Translators have a wide assortment of tools and systems at their disposal:

[1] Content can take many forms, such as marketing materials, business communication, software strings, online and print documentation, legal and regulatory materials, audio and video files, photographs, etc. For the purposes of this dissertation, "content" is understood as "any kind of audiovisual, visual, sound or textual information" (Mauthe and Thomas 2004:4). "Text," then is subsumed within the larger concept of "content" and will specifically refer to written digital text.

word processors; the Internet; computer-assisted translation tools; terminology management systems; content management systems; and cloud-based computing to name a few of the many technological offerings now available. Moreover, the previously-mentioned outsourcing model partners companies and language service providers that are located throughout the world, which in turn allows and requires language professionals to be adept at working off-site or remotely. As a result, translators often collaborate virtually with colleagues located elsewhere.

Each of these advances has changed how work is distributed and performed among virtual team members, the types of content that are translated, and the skill sets required to complete work. The expanded repertoire of resources changes the nature of translation in and of itself, and subsequently influences the nature and progress of the task both behaviorally and cognitively. Likewise, the translation task impacts the tools required of professional translators and language service providers, giving rise to new tools and processes to aid translation. These changes are reinforced by the fact that many companies require the use of translation tools to complete assignments, which changes the way translators approach their jobs. The development of these technologies, be they translation memories, machine translation, concordancers, alignment tools, or corpus-building tools, has largely responded to the needs of translators in an effort to support their work, to achieve productivity gains, and to address the ever-increasing volume of work and time pressures (Hutchins 1998; Austermühl 2001; Bowker 2002; Kenny 2011; Dunne 2013).

Translation memory (TM) tools, in particular, are used by translators working with non-literary texts. These tools present the translator with the source language (SL) text to be translated in segments, typically on a sentence-by-sentence basis, and prompt the translator to render a target language (TL) version. The SL and TL segments are then stored as an aligned pair in a database that allows subsequent reference, retrieval and reuse of these bilingual sentence pairs, or translation units. If a SL segment is subsequently encountered that is identical to one stored in the TM, then the translation is proposed to the translator who in turn determines if this match is appropriate in the new context. This proposed match is considered to be an exact or 100% match. Likewise, if a SL language segment is similar to that of a translation unit that is already stored in the TM, the stored translation is presented to the translator who can then choose to accept the translation as is, edit the translation to better convey the meaning of the new source language segment, or reject the proposed translation entirely and create a new one. These inexact, but similar, matches are called fuzzy matches.

Statement of the Problem

The question arises as to the impact that translation memory tools has on the translator, and his or her ability to complete work. These tools are often touted as a silver bullet to increased throughput and to overall ease of translation, but these claims have not been sufficiently scrutinized; the mutual effect of technology and translation process needs to be better understood. In particular, the notion that translation memory makes the translation process easier requires investigation, since the translation task itself has been changed. Instead of translating an entire text from scratch, the translator now completes a hybrid of translation, editing, and cross-language verification. While it is clear that the translation task has in fact changed and requires a different type of intervention by the translator, little empirical evidence is available to suggest that the task is easier or less effortful. In fact, research on effort in translation is largely absent from the literature. A more thorough review of effort and translation appears in Chapter 2, but the gap in the literature on effort and computer-assisted translation warrants further inquiry.

Beyond filling a noted gap in the literature, understanding the effort required of translators when working with translation memory tools has ramifications for the economics of translation. Current compensation models provide discounts for fuzzy and exact matches, while segments for which no corresponding translation is stored in the TM are billed at full price. This tiered pricing model implies that working with fuzzy and exact matches is less effortful than translating a text and is consequently paid at a lower rate. Nevertheless, empirical studies have not substantiated this assumption that is heavily relied on by translation buyers.

To date, only a handful of studies have explored the relationship between a translator’s cognitive effort and the use of translation memories, specifically when translators are presented with possible translations for SL segments. Research on post-editing of machine translation output provides a point of departure, including an in-depth discussion by Krings (2001), who argues that post-editing effort can be classified as temporal, technical, and cognitive. Several scholars have pursued further work with these divisions to explore post-editing of both machine and human output. For example, O'Brien (2005, 2006a, 2007) investigates the different types of effort that are exerted by translators and post-editors, and how effort and processing speed may differ depending on source-text features (called "negative translatability indicators" in the MT literature). Several other scholars have also looked at effort as it relates to post-editing of either human or machine translation output and have introduced novel metrics to approximate cognitive effort.[2] These studies address and debunk some of the commonly-accepted views of segmentation and translation memory, but have not yet tackled the question of the distribution of effort during the editing process. Further work is necessary to better understand effort during the translation and editing tasks, and the role that technology plays in changing the process.

Purpose of the Study

This dissertation focuses on cognitive effort during the translation process when translation memory is used. More specifically, two questions are addressed by means of an experimental study. The first question is whether the use of translation memory affects the cognitive effort of the translator during the process of translating segmented texts compared to translation without the use of a TM. This comparison is of segmented texts

---

[2] For example, Guerberof (2009) investigated the amount of effort required of translators when editing translations produced by human translators that had been stored in translation memory, and compared this effort to that needed to post-edit machine translation. Other metrics have been introduced in related tasks, such as sight translation, including sequences of repairs and pause position (Shreve, Lacruz, and Angelone 2011), and average pause ratio (Lacruz, Shreve, and Angelone 2012; Lacruz and Shreve 2014).

only, as the experimental design will use a segmented text to investigate this potential difference. One expected difference is that a greater amount of cognitive effort is required to edit fuzzy matches proposed by the TM than to produce a translation without the aid of a TM. Similarly, it is hypothesized that variation in effort will be observed with respect to the overall effort contour. In other words, cognitive effort may remain relatively constant regardless of whether a TM is used, however the application of effort in specific portions of the text or segment may differ. For example, the areas that show the greatest amount of effort are anticipated to change depending on if the translator is editing a proposed fuzzy match or if the translator is translating the segment from scratch.

The second research question to be addressed in this study is whether translators perceive translation memory proposals as useful to the translation task. Again, the goal is to investigate the level of cognitive effort needed to edit fuzzy matches; however, the variable under consideration is the perceived quality of the fuzzy or exact match presented. The scenario in which a translation memory tool proposes a poor quality fuzzy match, while undesirable in professional working conditions, is realistic, and one that a working translator can (and often does) face. The effect of poor TM quality on the translator's cognitive effort might manifest itself as an overall increase in effort when compared with the effort required to revise a good-quality TM proposal, to translate the segment from scratch, or both. Another possible effect would be a change in effort allocation or the effort contour resulting from a poor quality fuzzy match. Overall, the aim is to explore how a problematic fuzzy match from the translation memory will affect the cognitive effort required of the translator, and to identify the threshold at which the translator chooses to abandon editing the segment and instead re-translate the segment from scratch.

In sum, two research questions are addressed in this dissertation. The first question is whether the use of translation memory affects the translator's cognitive effort. The second question is whether the translator's perception of the effort required to complete a task coincide with the actual exerted effort. Both of these questions are experimentally investigated in an attempt to illuminate the effects resulting from the use of translation memory.

Hypotheses

As noted in the previous section, differences in cognitive effort allocation are expected between all three conditions; that is, when translating from scratch, with a proposed fuzzy match, or with an exact match. The first hypothesis is that cognitive effort is expected to be greatest when the translator is working with a proposed fuzzy match. This increased effort is hypothesized to arise from translators being required to identify deviations between the source text and the proposed translation, and then to edit the segment as necessary until reaching a final translation. Reviewing and verifying exact matches is predicted to be the least effortful, with effort when translating a segment in its entirety falling in the middle.

The participants' perception of the effort required to translate or edit each segment is expected to align with the number of changes introduced in the segment, but not necessarily with the actual effort exerted by participants. Results suggestive of a correlation between the number of changes introduced and the perceived effort would corroborate findings reported by O'Brien (2008).[3] Participant's ranking of overall ease to edit or translate each segment is expected to correlate with the process data obtained during the translation and editing task. Based on their experience, participants can be expected to be aware of the difficulty of introducing changes, rather than basing this measure solely on the number of keystrokes or mouse clicks required to modify the segment.

In short, translators are anticipated to exert the most effort when working with a fuzzy match proposal from the translation memory. Exact matches are expected to require the least amount of effort, while new translations fall between the two. Moreover, the allocation of effort of the participants is hypothesized to vary depending on the type of match that is proposed. Participants are likely to exert more effort prior to introducing any changes when presented with a fuzzy match or exact match, while segments that require a novel translation are expected to elicit more effort throughout the segment. Translators are also predicted to perceive the amount of effort required to translate a segment based on the number of changes that have to be introduced. The translator's perception and their behavior are hypothesized to align.

---

[3] O'Brien's study (2008) investigated cognitive processing of translation memory segments using eye tracking, measuring the effort exerted by participants using pupil dilation metrics and eye movements. Perceived effort was measured using a retrospective survey of each segment using a five-point Likert scale. A linear correlation was found between the fuzzy match percentage of each segment and the effort that was perceived; however, no mention is made of a correlation between perceived effort and online process data.

Significance of the Study

As has been previously discussed, the use of translation memory systems is growing in the language industry, and thus plays an increasingly important role in how translation work is divided, completed, and delivered. Translation software providers often cite the productivity gains obtained from using translation memories, machine translation, and terminology management.[4] These claims ostensibly demonstrate the potential effect of translation memory and of translation re-use on translation productivity. These gains in speed however, are not without their tradeoffs. DePalma, Stewart, and Whittaker (2010) note that translators are able to save time, but "it can be tedious to sort through a list of sentence candidates, select the most appropriate one, and then edit it to be consistent and correct in the current context" (21). This problem is further compounded if the translator does not have sufficient training or if the translation memory in use has not been properly managed.

These productivity gains have implications for the economics of translation. Translation memories provide translators with previously translated material that is to be reviewed, and that is often billed at a lower rate than new words to be translated. One of the driving forces behind this pricing structure is the expectation from translation buyers

[4] A quick review of the predominant translation memory tool providers' websites clearly reveals this type of verbiage. SDL (maker of Trados Studio), claims productivity gains of 40%, with the possibility of even greater speeds depending on the content being translated (http://www.sdl.com/products/sdl-trados-studio/). Likewise, Kilgray Translation Technologies (maker of memoQ) states that the software "optimizes the productivity of translation and localization" (http://kilgray.com/solutions/user-group/translators). Neither of the marketing claims specifies how these figures were reached or how these gains actually manifest; however, this type of claim is a typically-cited advantage of CAT tool use. [Both of the websites cited above were accessed 22 January 2013.]

for a discount on previously translated work (DePalma and Stewart 2012:17). Language service providers differ somewhat on how fuzzy matches are discounted or priced, but typically do so at a percentage of the new word rate. Repetitions are typically priced at 36.64% of the offered rate for novel words (ibid). This tiered pricing model should be a reflection of the amount of work or effort required by the translator to complete the work, given that the translator is working with a proposed translation.

The question arises, however, if the cognitive effort of the translator aligns with these pricing models and industry pressures. It may be the case that the amount of effort required on the part of the translator does not coincide with currently established rates for compensation, which in turn could undermine the pricing models employed. Dragsted (2008) explores how TM use may be misaligned with cognitive processes in translation. She hypothesizes that there is an effect on the translation process when using these TM systems since there is an unnatural focus on the microcontext, which does not necessarily coincide with the translator's mental representation of the text. Her study demonstrates that for professional translators, using TM systems has a "constraining effect on their cognitive behavior and mental representation of the text" (251) and ultimately changes the translation and revision tasks.

Likewise, should the amount of cognitive effort of revising proposed fuzzy matches in fact be greater than or similar to that of translating a text without a TM, it may be necessary to re-evaluate the per-word approach to pricing, and perhaps shift to a pricing model that takes into account the amount of effort required. Guerberof (2009) alludes to this issue when investigating post-editing of TM and MT output, and the effects of using either translation memory matches or machine translation proposals on productivity and quality. In this study, the researcher compares the amount of time required to edit the two outputs, and she concludes that machine translation output is processed faster than TM fuzzy matches. While the results of the study are not statistically significant, Guerberof does present several important ideas that warrant further investigation, particularly that a greater understanding of how tools affect the translation process will help prepare “translators and the translation community… [to] come into the negotiating arena with the knowledge necessary to reach common ground with translation buyers” (20). In this sense, research on how these tools affect cognitive effort will provide empirical data about relative effort, which can provide an informed basis on which decisions can be made concerning how translators should be remunerated.

Tool designers have an equal stake in understanding the cognitive effort expended by translators when using translation memories. While the usability of specific tool design and user support is outside the scope of this dissertation, a better understanding of how translators apply effort during the use of translation memory would be of particular import to tool designers. Proctor and Vu (2012) describe three actions that occur when a user interacts with a computer system: users must identify displayed information, select responses based on information, and execute those responses by entering commands (21). In order for this interaction to be efficient, the “interface must be designed in accordance with the user’s information-processing capabilities” (21). Several different kinds of effort could be involved in the user’s information-processing capabilities. Krings’ (2001) tripartite model of effort—effort is temporal, technical, or cognitive—is useful in classifying the type of effort being exerted. For example, the physical manipulation of the software would be technical effort. Mouse clicks, keystrokes, and other interaction with the system are measures that can help quantify and observe this type of effort. The time on task, in contrast, would be Krings' notion of temporal effort. This type of effort can also be observed by measuring pauses or the elapsed time between executing commands.

The third type of effort that Krings outlines, cognitive effort, though is not directly measurable. Instead, both technical and temporal efforts indirectly indicate the underlying cognitive effort required of the user. This dissertation focuses on cognitive effort, and uses keystroke logging, mouse clicks, and time stamps to approximate the exerted cognitive effort when using translation memory tools. The different measures of effort and the underlying cognitive effort would be useful to TM system developers in creating efficient and effective programs. By minimizing the amount of effort required to interact with the system and allowing translators to focus on their work, developers can help facilitate the translation process.

Outside of the realm of practicing translators, language service providers and translation buyers, differences in cognitive effort when translation memory is used have a direct impact on how translators should be trained. Few works discuss specifics of how translators should be trained in computer-assisted translation. Samson (2005) argues for a broad definition of computer-assisted translation, and suggests a cross-curricular approach that includes varied and practical exercises, in conjunction with the establishment of partnerships with key industry players. While these suggestions could be considered well-founded, a greater understanding of how translators interact with computer-assisted translation tools will allow for a more refined approach to training that is informed by empirical research and that could better align with research on pedagogy best practices.

Overview of the Dissertation

This dissertation is divided into six chapters. The present chapter introduces the purpose of the study and justifies its significance. Chapter 2 lays the conceptual framework for the study, bringing together the literature on translation technology, post-editing, working memory, and effort in order to design the experiment that is described in Chapter 3, and interpret the results that are presented in Chapter 4. As mentioned, the third chapter addresses methodological considerations taken into account when designing the study, and introduces a novel way to collect translation process data. The chapter also sketches a profile of the participants included in the study. Chapters 4 and 5 are closely related, in that the former reports the results of the experiment, while the latter aims to contextualize the findings. The amount of data generated by process-oriented research can be expansive, and consequently, interpreting the data and drawing valid, salient conclusions can prove challenging. Nevertheless, Chapter 5 aims to elucidate how the data obtained from the experiment may lend support to the hypotheses, and raises further questions to be studied at a later time. Chapter 6 synthesizes the findings, and contextualizes the results within the larger translation process. Moreover, we suggest implications of the study, particularly within the context of translation pedagogy, translation tool design, and the economics of translation. Special mention is also made of the research design, and how novel data collection methods can increase the pool of participants in empirical, process-oriented experiments, while still employing established metrics to measure cognitive effort.

CHAPTER 2

Conceptual Framework

Given the interdisciplinary nature of translation studies, scholarship from a number of areas must be reviewed to address the research questions outlined in Chapter 1. To do so, the overarching topics that are involved in this dissertation will be reviewed in turn. First, the state of the art of translation technology will be discussed, and some of the major types of computer-assisted translation tools are mentioned. The role of technology in the work of the translator will be contextualized, as will its use throughout the translation workflow and the ways in which translators have implemented these tools in an attempt to achieve efficiency and productivity gains. The subsection on the role of translation technology will conclude by focusing specifically on translation memory, describing how these systems work and the role segmentation plays in their use.

Next, the literature on post-editing of machine translation will be explored, starting with a brief mention of the types and levels of machine translation post-editing. The relationship between post-editing of machine translation and the editing of human output will then be established. This relationship is paramount, as it grounds this dissertation's research that aims to measure the level of intervention of the translator.

Brief mention will be made of current models of working memory to segue into a discussion of effort. Krings' (2001) model of effort and his division of effort into temporal, technical, and cognitive effort will be used to frame this discussion and its investigation within the realm of translation studies. Particular emphasis will be placed on recent work related to cognitive effort and its relationship to translators using technology in their daily work.

Translation Technology

Translation software and technology used by translators to aid their daily work can be examined from several angles (Alcina 2008); however the state of the art will be described here by classifying these systems based on function and on the level of user intervention. In the context of this dissertation, user intervention is understood as the involvement of the human translator in the translation process. This definition is based on Quah's (2006) discussion of machine translation, which details the typical lack of any reference to human involvement in the machine translation process. Quah laments this absence of human involvement in most machine translation descriptions, and cites several scholars (e.g., Balkan 1992; Archer 2002) who note the ambiguity introduced by the term *human intervention*. Given the lacuna of a more appropriate term, though, it persists in the literature.

The range of tools available to translators has expanded substantially from the initial systems developed in the 1950s and 1960s and those described by the Automatic

Language Processing Advisory Committee in 1966, and now include terminology management systems, corpus tools, translation memory, and machine translation (Hutchins 1998; Austermühl 2001; Bowker 2002; Kenny 2011; Dunne 2013). While terminology management systems, term extraction tools, and corpus tools arguably fall outside the scope of this dissertation, brief mention is made to demonstrate the various types of support available to translators. Future avenues of investigation could explore how each of these tools can impact the translation process. Moreover, these additional tools are often used in conjunction with translation memory, and additional research is necessary to understand the impact of the simultaneous use of multiple tools.

Terminology Management Systems

Terminology management systems first appeared in the 1960s to support efforts to develop fully-automated machine translation systems, and have relied on the documentation of concepts in specific domains in order to support the creation of multilingual content. These systems can be defined as "software products (programs and program packages) that are designed for the management of terminological data. They enable the user to collect, store, manipulate, and retrieve terminology" (Schmitz 2001: 539). Initially, these systems were a response to rapidly changing terminology in specialized domains and were integrated into machine translation systems in an attempt to improve machine-generated output (Hutchins 1998). Machine translation will be discussed in greater detail in a later subsection, but rule-based machine translation

(RBMT) systems are based on a three step process and rely on MT system users to document language features, syntax, and terminology in order for the MT system to render a translation (Hartley 2009). The first stage is an initial analysis, in which the MT system analyzes the source text based on a pre-determined set of rules, which may take into account part-of-speech categories, verbal tenses, aspect and modality. Next, the transfer stage, "relies on a bilingual dictionary and mappings between the abstract structure describing an SL sentence and a structure underlying the corresponding target language (TL) sentence" (Hartley 2009:122). Here, managed terminology can greatly support the automated translation system, as lexical items that have been stored in the source language automatically trigger target language equivalents. The final generation stage creates the target text as a "grammatically correct sequence of TL words" (ibid).

While initial efforts to develop machine translation failed to achieve complete automation at the envisioned level of quality, the importance of terminology management has remained. The previously described RBMT example outlines a more lexicographic approach to managing terminology, in that source and target language term pairs are stored. In the intervening years, terminology management has shifted from a lexicographical documentation to a concept-oriented one. These systems have grown in complexity, and now integrate into controlled authoring environments as well as knowledge and content management systems. International standards bodies have developed best practices to guide terminology work (e.g., ISO 704:2009), and scholars have argued that systematic documentation allows users to gain greater insight into a particular domain more quickly than through an alphabetical listing of terms (Wright and

Budin 1997:326). Moreover, the type of information documented and, in particular, data categories related to terminology work have been standardized (ISO 12620:1999) in an attempt to improve interoperability between terminology management systems. When used in conjunction with other translation technologies, terminology management systems can provide multilingual terminological support to the translator and help maintain consistency across documents and digital content.

Terminology Extraction Tools

Related to terminology management systems, terminology extraction tools allow terminologists or translators to automatically process digital texts to identify potential term candidates and their corresponding contexts. (Kenny 2001; Ahmad and Rogers 2001). The extraction algorithms are often based on word frequency and allow translators to vet term candidates before populating a termbase. Some extraction tools also allow parallel source and target texts to be processed in an attempt to identify both source and target language term candidates. These term candidates, however, are not necessarily all of the terms that are present in the text, nor is frequency the definitive way to determine whether a lexical unit is in fact a term. Translators or terminologists using these tools ultimately must decide whether the identified lexical units are in fact terms, and thus should be stored in a termbase, or if the proposed terms should be discarded. Nevertheless, these tools allow translators and terminologists to quickly identify a large number of terms and, ideally, document these terms prior to translation. Moreover, term extraction tools enable the user to capture not just the terms, but also the sentences in which they appear as contexts or examples of authentic usage. How widespread the use of these tools is debatable, since terminology management is not always performed in translation projects or in authoring, much to the chagrin of translators who are then left to untangle conceptual inconsistencies.

Corpus Tools

Corpus tools are more common in current versions of commercial software available to translators than in previously available tools, yet the full potential of corpus tool integration with terminology management, translation memory, and machine translation systems has arguably not yet been realized. Corpus tools allow translators to collect texts or content in the source or target languages, or both, and then make these texts available for reference. Users can generate word lists or concordances, and can review these texts and lexical items in context as a translation aid (Bowker and Pearson 2002). A number of benefits of corpus analysis tools have been discussed by scholars, particularly for training and professional purposes, and several computer-assisted translation tools companies have attempted to integrate corpus analysis tools or features into their products.[5]

---

[5] For a thorough discussion of corpus analysis tools and how these can be used for both research and professional purposes, see Bowker and Pearson (2002), Zanettin (2002), and Olohan (2004). A general review of corpus tools and their introduction in translation studies is also provided by Bowker (2002) and

As an example of how these systems have been developed thus far, consider the case of SDL, one of the leading translation software companies in the industry today. The computer-assisted translation software developed by SDL, Trados Studio, is not marketed or touted as a corpus tool, but includes several corpus-related features, including project reference documents, PerfectMatch technology, and AutoSuggest dictionaries. While this implementation of corpus tools is not the only approach that has been taken thus far in the language industry, it is indicative of a number of ways corpus tools have been integrated into computer-assisted translation software. As their name suggests, reference documents are texts that are stored with the project files and that the translator can consult as needed. Reference documents are not interactive, but can be opened, viewed, and searched when the translator wants to view a comparable text. This rudimentary approach of providing a corpus of related texts does not provide much real-time support during the translation task. Translators benefit from these documents, though, as they prepare to translate. Moreover, translators and clients can store additional reference materials related to the project for future use.

The second way corpus tools have been integrated into SDL Trados Studio is through PerfectMatch and AutoSuggest dictionaries. The idea of PerfectMatch, according to SDL, is a comparison of "updated source files to a corresponding set of existing bilingual documents rather than to a translation memory" (SDL 2013). Here, the software attempts to leverage previously translated documents or files to support current

---

Kenny (2011). Buendía-Castro and López-Rodríguez (2013) approach the use of corpora from a slightly different perspective, and provide two corpus-based approaches to training translators.

translation projects. To do so, the system relies on entire texts rather than on a sentential or sub-sentential segmentation. SDL Trados Studio compares the text to be translated with the previously translated source text and identifies larger passages in the text to be translated that are identical to the previously translated text. Then, the software inserts the corresponding translation for this larger section. Implementing corpus features in this way allows language service providers to reuse larger portions of previously translated material. In doing so, translators can then more quickly identify and focus on content that has changed from one version to the next. A drawback to this approach, however, is that its use is predicated on having previously translated material in a proprietary SDL file format that is only created when translating documents using SDL Trados Studio. To take advantage of this feature, additional upfront processing of the texts is required, which is labor-intensive and time-consuming, ultimately hindering widespread use of this corpus feature.

To support sub-sentential segmentation and to leverage previous translations stored in translation memories, SDL uses a bilingual corpus to generate AutoSuggest dictionaries.[6] Far from a traditional lexicographic dictionary, these language resources are created when the software analyzes large quantities of parallel corpus data in the form of translation memories to create approximate sub-sentential translation proposals. The translator is dynamically presented with these translation proposals as he or she moves through the text and can choose to use or discard the suggested text. In this instance, the

[6] Translation memories are a type of corpus that is bilingual and parallel (Olohan 2004: 176). Further discussion of translation memories appears in a subsequent section, *Translation Memory Systems and Segmentation*; however, the use of this type of corpus is necessary when creating sub-sentential matches.

aim is to identify repetitive phrases or turns of phrase that can improve the translator's ability to produce a translation that is similar in style and tone to previously translated material. Likewise, the translator can improve his or her efficiency. The corpus itself is not referenced directly by the translator, but rather the corpus provides the input for a resource that can be leveraged during translation.

This extended example of SDL's flagship product for freelance translators and language service providers is solely illustrative, and should not be interpreted as the only way that corpus-related features can be integrated into computer-assisted translation tools. In fact, corpus tools like WordSmith tools, ParaConc, and MonoConc that have been developed for corpus research and the creation of corpora for specific purposes can also be used in conjunction with computer-assisted translation tools to better effect. Specifically, translators can benefit from collocation and colligation queries, word and frequency lists, and lexical item distributions that the previously mentioned corpus tools can provide. Nevertheless, these tools are frequently unknown to practicing translators and little research has been conducted on practicing translators' use of these systems (Zanettin 2002; Olohan 2004).

Translation Memory Systems and Segmentation

The translation tools described thus far — terminology management systems, terminology extraction tools, and corpus tools — are used to support the translator in the translation task. These tools are not designed to specifically translate a text or to provide an environment in which a translator can render a target language text, but instead are used to aid translators once they begin to translate.[7]

Translation memory (TM) systems, in contrast, often provide a standalone program that allows translators to segment a source text and then translate segment by segment. Older TM systems, such SDL's Translator Workbench, integrate into authoring tools and allow the translator to work directly in the source text. Translators benefit from this integration since they do not have to learn a new piece of software and can use authoring tools with which they are familiar. The tradeoff, however, is the inherent risk of corrupting the source file. In contrast standalone programs allow translators to protect source files by extracting the translatable content into an intermediary file; in this new file, translators then render their translations. Once completed, the translated content is reintegrated into the original file type to generate a target language file.

The previous example, SDL Trados Studio, is a standalone translation memory system. Hutchins (1998) describes the development of the translator workstations or workbenches, and asserts that they are an outgrowth of initial attempts at fully-automated machine translation. These workbenches often integrate with the previously-mentioned translation tools — terminology management systems, terminology extraction tools, and corpus tools. Unlike in machine translation, however, the translator ultimately oversees

[7] It should be noted that while SDL Trados Studio incorporates corpus-based features, it is not designed as a corpus tool. Likewise, a terminology management system such as SDL Multiterm can provide terminological information to translators as they work, but Trados Studio itself is not a terminology management system.

and controls the creation of target language material and is responsible for providing a translation of any content that is not stored in the translation memory.

As mentioned above, translators use TM systems to translate a source text and recall previously translated material that has been stored in a linguistic database (Bowker 2002). In the absence of these systems, translators are required to render a target language version of every sentence in the text, even if the text is repetitive or has been previously translated. Occasionally, a previous translation is available to translators as a reference, although "it may prove difficult and time consuming to locate the necessary segment" (Bowker 2002: 93). Consequently, translators spend considerable time "searching, copying and pasting, and editing, [when] it would have been faster to retranslate the segments from scratch" (ibid: 94).

TM systems aim to solve this challenge by automating the retrieval of previous translations. To do so, the tools must be able to analyze and to store discrete portions of the text and their translations. Texts are typically segmented at the sentential level, although translators and language service providers can choose to segment the text at the sub-sentential or paragraph level (Quah 2006: 100). Translators then render a target language version of each source language segment. Each source text and translation pair is called a translation unit, which can be referenced or reused later in the text, in the project (if the project encompasses multiple texts), or in subsequent projects.

As the translator progresses through the text, the TM tool compares each source language segment with the previously stored translation units. If a source segment is identical to a previously stored translation unit, the tool proposes the corresponding translation to the translator who determines if the rendering is appropriate. This type of match is an exact match or a 100% match. If a segment is similar, but is not an exact match, then the stored translation is displayed to the translator who can then accept, edit, or reject the proposed translation. These inexact, but similar, matches are called fuzzy matches.[8]

Segmentation deserves special mention at this point, since translation as described in the preceding paragraphs shows a marked shift in how the translation task is carried out, in that the text is presented, and subsequently stored, in a decontextualized, fragmented manner. Without the use of a translation memory, translators render the target language version from scratch. When using the translation memory tools, translators work with proposed translations provided by the tool in a segmented text. According to Quah (2006), segmentation performed by translation tools lends itself more naturally to Latin alphabet-based languages, which divide lexical items and sentences with spaces and punctuation marks, while Asian languages require a different type of intervention to leverage previous translations. While this may be true in the sense that character-based comparisons are possible, the notion of a text being translated segment by segment can prove problematic within these same script languages if we consider the rhetorical structure of a text and the order in which information is presented. How a text is structured may be dependent on the language in which it is written, so a parallel, one-to-

[8] The determination of fuzzy match is dependent on the algorithms used by each system. Bowker (2002) describes character-based matching for translation memories, and this is one of the main ways that mainstream TM tools determine matching statistics. Quah (2006) outlines several parameters that may govern the threshold at which segments of a certain match value are displayed.

one correspondence of the sequence of source- and target-language segments of a text could in fact prove counter-productive. Esselink (2000: 367) notes this disadvantage, stating "translators do not have the opportunity to change the overall structure of the text, i.e., to change the sequence of sentences within a paragraph." As a result, translating a segmented text using translation memory may only be suitable for specific text types or genres, or may require additional work to create an acceptable rendition. For example, writers of argumentative texts in English may favor a linear progression of ideas and less flowery language, while Spanish writers of the same text type may prefer a more circular progression and a greater number of rhetorical flourishes.

Machine Translation

Thus far, some discussion has been made of machine translation (MT) as it relates to terminology management. Machine translation as a tool used in translation workflows, though, merits further discussion. Arnold et al. (1994: 1) define machine translation as "the attempt to automate all, or part of the process of translating from one human language to another." This definition highlights a key difference from translation memory and human translation: automation. While translation memory systems place the translator squarely in control of the translation process, machine translation systems leave the translation to the system's algorithms.

There are several types of MT systems that users may employ to translate source language texts, and two will be mentioned here: rule-based machine translation (RBMT) and statistical machine translation (SMT). As noted in the terminology management subsection, RBMT is based on linguistic rules that have been created by the MT system developer. These rules essentially map source language syntax and lexicon with target language equivalents to render a TL version. Documented terminology can also be incorporated to improve target language renditions of the source text. In contrast statistical machine translation uses bilingual parallel corpora as input (Arnold 1994; Ping 2009). These systems analyze the aligned corpus to calculate frequencies, which can then be used to model a correspondence between source language words and phrases with target language counterparts. Newer systems now combine several ways to generate target language output, and these hybrid systems continue to show promising results (Hutchins 2010).

Role of Translation Technology

The number of different translation technologies at the disposal of the translator suggests several ways to carry out a translation project. International standards bodies have proposed a recommended workflow for translation (i.e., ASTM International 2006), and note that how a project is executed is (or should be) determined by the project specifications. Several steps are commonplace, with terminology management ideally occurring before translation, and editing and proofreading necessarily following translation. At every stage, the previously cited translation tools can be used to improve efficiency and consistency of the translator or translation team. Many of these tools have been developed in response to specific needs arising in these translation workflows. For example, terminology management allows stored multilingual terminological assets to be quickly accessed during the translation process, rather than requiring repeated searches for appropriate foreign language equivalents. Terminology extraction tools enable an initial list of terminology to be compiled to avoid manual term identification. Each of these tools represents an attempt to facilitate the process and make the work of the translator more efficient.

Another consideration is that many translators are subcontractors, and they may only be involved in one particular stage of a particular project. As a result, the use of these technologies must be coordinated so that content can be divided and shared among a team of language service providers. This coordination further illustrates how technology is an integral component of the translator's task and shapes how translation agencies and companies distribute work. Moreover, translation agencies often require translators to be versed in using these tools and to be able to support the technological complexity of the projects that are undertaken. These technologies also regularly evolve, requiring the translator to be able to learn new tools and stay abreast of changes in the field.

Post-editing

As described in the previous section on machine translation, advances in MT and the quality of MT output has led to its increased use and a growing need for post-editing of machine translation output. While machine translation systems have been under development since the 1940s and 1950s (Hutchins 1998), the market for machine translation (MT) has been limited to specific domains and relatively few translators have experience post-editing MT output. Now, a greater diversity in MT tools, increased quality of MT output, and large volumes of text that require translation, have helped secure MT a more central role in large-scale translation projects. As a result, freelance translators are now learning how to post-edit and more LSPs are beginning to offer post-editing as a service.

Types and Levels of Post-editing of Machine Translation

When post-editing machine translation, companies and post-editors must determine the level of intervention required to bring the text to the requisite level of quality. This standard is variable, since the purpose of machine translation may differ depending on the goal(s) of the project and the client's specifications. Guidelines for post-editing are not particularly standardized either; however two are worth mentioning here. Both have been proposed by industry professionals, and highlight the importance of identifying user needs and show an awareness of the strengths and weaknesses of machine translation.

The first set of guidelines to be discussed is proposed by Spalink, Levy, and Merrill (1997), who classify post-editing into three levels: editing for confirmation, for comprehension, and for publication. These post-editing levels center on the end user's need for the machine translated content. The first, editing for confirmation, is to determine whether a particular topic has been covered in another language, and the authors describe this form as the lightest level of post-editing. One scenario proposed as appropriate for this level of editing is in journalism, to determine if similar discussions on a particular topic are happening in another language. The proposition for this type of post-editing is that “details of the discussion and the form of the MT output are irrelevant” (3) and corrections are only made related to specific identifiers established prior to the translation.

The second level of post-editing in the scheme proposed by Spalink, Levy, and Merrill is editing for comprehension, which aims to make the content understandable to the end user. When working at this level of intervention, the post-editor corrects any errors that “change meaning” (66). These changes require more time than post-editing for confirmation, but the goal is not to create a translation which rivals that of a human translator. This type of post-editing of machine translation might be appropriate after the end user has determined that foreign language content is adequately covering the topic in which the end user is interested (such as the case for post-editing for confirmation).

To achieve a publication-ready text, the greatest intervention on the part of the post-editor is what has been termed “editing for publication.” As the name suggests, this type of intervention requires any and all errors to be corrected, and the text must be ready to be distributed in any number of publications. Spalink, Levy, and Merrill (1997: 4) note that given “the nature of machine translation, there are many cases in which machine translation is not most efficient for publication." Nevertheless, at the time of their writing, this level of post-editing was still imagined and undertaken in specific instances.

Having reviewed post-editing guidelines from almost two decades ago, it is surprising to see how little has changed in the intervening years. In 2004 the Translation Automation User Society (TAUS) was established as a clearinghouse for resources related to machine translation. In subsequent years, the mission of TAUS has expanded to encompass the creation of guidelines for MT use and of implementation strategies. One guideline proposed by TAUS in 2010 is related to post-editing of machine translation. Instead of three levels, TAUS (2010) proposes only two: full and light post-editing. The former aims to approximate quality "high-quality human translation and revision" while the latter strives for quality that is, in TAUS' terms, "good enough." Full post-editing seems to align with Spalink, Levy, and Merrill's notion of post-editing for publication, while light post-editing approximates post-editing for comprehension.[9] The idea of editing for confirmation seems to disappear in the TAUS recommendations; however, this may be in light of advances in machine translation that obviate the need for this level of editing since the raw output from these systems has improved in recent years.

Editing of Human Translation vs. Post-editing of Machine Translation

---

[9] TAUS (2010) provides specific guidelines that define what each type of post-editing entails. The organization is quick to point out, however, that starting with poor MT output can ultimately affect the level of quality can be achieved of either type of editing, and that more extensive intervention may be needed in such cases.

The previously discussed guidelines implicitly suggest a similarity between post-editing of machine translation and editing of human translation. As has been noted, the description of full post-editing by TAUS specifically outlines the goal of this level as being on par with “high quality human translation and revision.” Indeed, both editing of human translation and post-editing of machine translation follow the same basic process. In both cases, a translator or editor compares the source text with the proposed target text, and makes a determination as to its quality. The editor also identifies divergences in the content of the source and target segments, and takes any corrective action necessary to bring the target language segment in line with the source text. This cross-language evaluation is required in both human translation editing and machine translation post-editing. The narrowing between human and machine translation may at first seem indicative that the tasks are one and the same and that research on one type of editing would describe the other. Nevertheless, research on post-editing of machine translation cannot be applied directly to the editing of human translation without it first being contextualized. For example, the types of texts that are being machine translated may differ from texts that are translated by human beings. The MT guidelines proposed by Spalink, Levy, and Merrill and the guidelines proposed by TAUS both emphasize this point and suggest that human translation is more appropriate in certain contexts. In fact, tools are available to determine the appropriateness of machine translation for a given text (Underwood and Jongejan 2001). Industry stakeholders also clearly distinguish in case studies between post-editing and human translation (i.e., Beaton and Contreras 2010).

Research on post-editing of machine translation, though, can be fruitful for examining the editing of human translation that uses a translation memory, particularly within the context of computer-assisted translation. The segmented nature of the task is similar since the segmented source and target texts must be compared. Likewise, certain features of source texts also pose challenges to MT systems. These features are termed negative translatability indicators, or NTIs (Underwood and Jongejan 2001; O'Brien 2005). An NTI can be a "linguistic feature, either stylistic or grammatical, that is known to be problematic for MT" (O'Brien 2005: 38). NTIs may include features like terminology, polysemy, nominal compounds, punctuation problems, and abbreviations, and often result in erroneous or flawed MT output, which in turn poses challenges to the post-editor. It should be noted, however, that the difficulty lies in the transfer from SL to TL.

A parallel can be drawn to human translation, in that certain source-text features pose problems to translators during the translation itself. Difficulty for human translators may occur during text comprehension, translation (or transfer) from SL to TL, or target text production. Arguably, these features vary across language pairs, text types, genres, and even individuals. That aside, certain source-text features are often identified anecdotally by translators and trainers as being particularly problematic to render in another language. A cursory review of language-specific educational sessions at several translation conferences highlights some of these potentially troublesome features, such as culturally-bound terminology, sexist language, punctuation, SL interference, proverbs, cognates and false cognates, capitalization, stylistics and grammar. For example, the most recent American Translators Association annual conference (2013) included sessions entitled: English-to-Spanish legal translation: Pitfalls and precautions; 100 most difficult words to translate into Portuguese; The return of false cognates and other fine points of Russian-to-English translation.

Beyond textual features that complicate the SL-to-TL transfer, the requirement that translators use machine translation and computer-assisted translation to complete their task allows for a comparison to be made between the two, in that these tasks are similar. Newer computer-assisted translation tools now integrate two types of proposed translations to translators as they work: TM proposals and MT output. The tools allow for editors to see matches, if available, that have been stored in a translation memory by translators. When no match is available, the tools will then provide a machine translation of the source segment to the translator. As a result, editors and post-editors now work with output from both MT and TM. Translators and editors are tasked with evaluating proposed translations and introducing changes where necessary. Moreover, editors are working with segmented texts, and must split their attention between the source and target languages. Research on post-editing may be appropriate in this context, particularly related to the influence of segmentation and technology on cognitive effort.

Working Memory

The previous discussion of machine translation post-editing and human translation editing highlights the type of intervention required of the editor. When working with either TM or MT translation proposals, the editor must read and understand the source text, evaluate the proposed translation, and identify any corrections that need to be made. All of this processing occurs in the editor's working memory. Since the purpose of this dissertation is to investigate the cognitive effort required when editing TM output, it is necessary that we examine working memory.

Several models of working memory describe cognitive effort. Three contemporary models tend to be referenced most frequently, namely Cowan (1995); Baddeley and Hitch (1974, 1976); and Ericsson and Kintsch (1995). In Cowan's (1995) model, working memory forms part of long-term memory (LTM) stores that are activated and the focus of attention is placed on up to four of these LTM stores at any given time. Baddeley and Hitch's model, in contrast, views working memory as a temporary storage space for information that is being processed and is only accessible for a short period of time. This model accounts for a relationship between working memory and long term memory; however, it is not subsumed under the same structure. The third model, that of Ericsson and Kintsch (1995), extends the ideas of Baddeley and Hitch by integrating literature on skilled memory in an attempt "to account for the large demands on working memory during text comprehension and expert performance" (211), and the fact that skills tasks can be interrupted and resumed without major effects on performance. Moreover, the Ericsson and Kintsch model proposes long-term memory retrieval via retrieval cues that are stored in short-term memory.

For the purposes of this dissertation, Baddeley and Hitch's (1974, 1976) model will be used for several reasons. Firstly, this model, as Shreve and Diamond (1997) note, is one of the most commonly accepted models of working memory outside of cognitive sciences. Second, this model has been integrated into several translation and interpreting models and working within the same framework will better allow for results to be compared. Moreover, Ericsson and Kintsch's model draws on literature from skilled memory theory and expertise studies, which may not be appropriate given the proposed experience of the participants, in that participants will presumably not have amassed the requisite 10,000 hours of deliberate practice that is often cited as a prerequisite of expertise. This model warrants further investigation in the field of translation studies, and future research may benefit from its use. Nevertheless, for this study, the Baddeley and Hitch model will be used in an attempt to describe cognitive processes during translation.

Baddeley and Hitch (1974, 1976) describe memory as having a visual-spatial sketchpad, phonological loop, central executive, and a long-term memory store. Of these subcomponents, the central executive is of particular importance when it comes to understanding cognitive effort, as it here that cognitive resources are allocated to the task or tasks at hand. These resources are finite and the amount of resources in use is typically referred to as cognitive load. Sweller's (1988, 2007) work on problem-solving is an outgrowth of, and partial response to, Baddeley and Hitch's model, and he defines three types of cognitive load: germane, intrinsic, and extraneous. Whether the cognitive load results in changes in long-term memory (germane), is the result of the "intrinsic complexity" of the task at hand (intrinsic), or the result of inappropriate instructional design (extraneous), the "sources of cognitive load are additive and cannot exceed the available capacity of working memory" (Sweller 2007:374). The limited nature of working memory is paramount to understanding cognitive effort, in that the amount of cognitive resources allocated to a particular task has a direct bearing on the remaining capacity of working memory. Cognitive effort, then, is the cognitive load required of the translator to complete a particular task.

Cognitive Effort in Translation

Shreve and Diamond (1997) take up the specific issue of effort in translation and interpreting, and highlight a number of ways to identify effortful processing, including think-aloud protocols and retrospective verbalizations (e.g., House 2001), a decrease in task efficiency, or the diminished ability to perform concurrent tasks. Similarly, effort has also been measured using fixation measures, such as duration in eye-tracking studies and pauses in keystroke logging methodologies. Eye-tracking studies have provided insight into segmentation of text by translators (Alves et al. 2010), cognitive effort and processing speed (O’Brien 2006a), and the ways in which translators allocate time to the translation and revision tasks (Alves, Pagano, and da Silva 2009). Likewise, studies using keystroke logging have investigated text segmentation (Alves et al. 2010; Dragsted 2004, 2005, 2008) and how difficulty, as observed by the translators, affects production speed, text segmentation, and effort to render a target language version of the text (Dragsted 2005; Dragsted and Hansen 2008).

Much more could be said concerning the measures of effort, and Chapter 3 will outline those measures specific to this dissertation. Brief mention will be made here, though, to Krings' (2001) triadic notion of effort. Krings divides effort into temporal, cognitive, and technical effort in order to differentiate the types of effort involved in the post-editing task. Temporal effort is the time required to post-edit, while technical effort is the physical manipulation of a keyboard and mouse via keystrokes and clicks to post-edit the text. Cognitive effort, then, is the remainder of the effort required to identify and correct errors in the text. He suggests that in many instances that this type of effort is not evenly distributed, as a less cognitively demanding change could in fact require a significant number of keystrokes or clicks to correct the segment. Likewise, a slow typist may require more time, but not necessarily more effort.

This division of effort is important to keep in mind when reviewing the literature, and has been cited in a number of the previously mentioned studies. Moreover, this division makes online process data of a translator's work essential to more closely assess the cognitive effort required during a given task. Taken separately, each division only provides a partial view of the effort exerted by the translator; taken together, a more complete view is possible.

Perception of Effort

Effort, as shown thus far, is only part of the overall picture and how the translator perceives effort is of equal importance. Anecdotal evidence provided by translators places the perception of effort at the forefront of decision-making for training, continuing education workshops, and conference presentations. Empirical research concerning how translators perceive the amount of effort required to complete their work is relatively scant, but Koponen (2012) provides insight as to whether post-editing operations align with the perceived effort of the post-editing task. Koponen's findings suggest that segment length may have an effect on the overall amount of perceived effort that is required to post-edit the segment. Moreover, post-editing operations that require reordering of the sentence seem to indicate a greater amount of cognitive effort, while participants exhibited less cognitive effort when making lexical changes. Additional work is needed to corroborate these findings, and to see if the perception of effort is similar when working with a translation memory.

CHAPTER 3

Methodology

The literature review highlights the need for an experimental study that investigates the research questions posed in Chapter 1. As discussed in the previous chapter, editing of human translation that used translation memory requires translators to read the source and target texts, evaluate the proposed translation, and determine any changes that may need to be implemented. The researcher is interested in seeing how the use of translation memory ultimately affects the effort exerted in the translation task using translation memory. Consequently, cognitive effort is the study's main object of investigation. Nevertheless, measuring effort proves challenging for several reasons. First, the empirical design of the study must be grounded in previous work on cognitive effort and computer-assisted translation. This foundation is paramount to ensure that the hypothesized effects on cognitive effort can be tested. Second, great variation in the values of participant variables can be challenging and could potentially muddle the obtained data. Variation in the participant pool is inevitable; however, care must be taken to minimize the number of variables that could be introduced.

In this chapter, the experimental design of the study is outlined, the data collection method is described, and the rationale for using keystroke logging and pause metrics is proposed. The chapter also establishes the background for the participants' perception of exerted effort. Participant selection is described in greater detail, as are the participants themselves, along with the justification for selecting specific participant variables.

Experimental Design

An experiment that allows for the collection of online[10] process data was chosen as the best way to measure the translator's cognitive effort during the translation process. An evaluation of the final product alone would not provide sufficient information to understand the cognitive effort exerted by each participant. For example, we could adopt Campbell's (2000) approach to product research and use choice network analysis to determine where translation decisions were made. Points in the text where there is wider variation among the participants' translations might suggest textual features that require greater cognitive effort. This comparison may prove fruitful in future work to isolate specific features and determine their effect on cognitive effort; however, for the purposes

[10] "Online" here is borrowed from psycholinguistics to refer to real-time data collection of the translation and editing process, rather than an evaluation of the final product. The study itself was also conducted over the Internet, which will be described in the subsequent subsections; however, in this instance, "online" should be understood as real-time process data.

of this dissertation, choice network analysis does not allow for cognitive effort to be measured in detail for each proposed type of translation memory match.

To obtain sufficient information to address the research questions, we adopt metrics previously used in translation studies that measure effort. As mentioned previously, Krings' (2001) model of effort provides a framework within which this discussion can be contextualized. Technical effort indicators, in the form of keystroke logging and mouse clicks, and temporal effort indicators, in the form of time on task and pause metrics, are both used to observe the translator's cognitive effort. These indicators are measured in an experiment conducted over the Internet that allowed for specific events in the translation process to be measured.

The study necessarily involved the participation of human subjects, and as a result, approval from Kent State University's Institutional Review Board was required. All of the necessary IRB documentation and forms, instruments required for data collection, recruitment scripts, and consent forms were developed prior to conducting the study. A Level-2 protocol (#13-293) was submitted for review on June 12, 2013. Approval was granted on June 17, 2013. All of the instruments, recruitment scripts, and approvals appear in the appendices.

Data Collection

In order to broaden the pool of potential participants, the experiment was conducted over the Internet. Collecting process research data via the Internet is relatively new in the field of translation studies, but this collection method is beneficial for a number of reasons. Online data collection expands participant pools to include translators from all over the world. Empirical process-oriented studies in translation to date have been site-bound, and thus are often limited in size and scope given the relatively small number of participants available to come into a laboratory setting. Limited participant pools may also require researchers to avoid limiting participant variables in order to ensure there are enough people available to conduct the study and to have enough power in their research designs. Consequently, results can be difficult to interpret and generalize to larger groups of translators since the observed effects may not necessarily describe specific populations. That is not to say that research conducted thus far with smaller participant pools and more general participant variables is not valid; quite the opposite is true. The inclusion of online data collection in the researcher's toolkit allows for researchers to be more specific in their selection of participant variables. Researchers can target more specific variables as they are not confined to the geographic location of participants.

To conduct this type of online data collection, researchers must adapt traditional process research designs to be effective in this virtual environment. Keystroke logging, for example, is often conducted using Translog (Jakobsen 1999, 2011). This easy-to-use software pioneered a number of process-oriented studies and continues to enable the collection of keystroke logging data for any number of studies. The drawback, however, is that Translog is limited to installation and use on a local machine. As a result, this highly successful tool cannot be used for Web-based data collection.

To address this issue, several researchers have developed their own tools that allow web server-based data collection. Alves and Vale (2009) developed CORPRAT as a way of querying larger data sets to investigate translation units while being able to capture keystroke logging data. The CASMACAT project (2011–2014), currently underway at the Center for Research and Innovation in Translation and Translation Technology (CRITT) in Copenhagen, provides an online workbench that allows participants to work remotely on the project and which can capture several types of data at a time. This homegrown approach to data collection has its advantages, particularly in that the systems can be customized to address the specific research questions that have been posed. Designing a system, though, requires a level of technical ability that many researchers lack. Consequently, finding ways to use or adapt currently available systems is one way to circumvent this challenge.

For this dissertation's study, Denkowski and Lavie's (2012) data collection tool, TransCenter, was adapted to investigate cognitive effort during computer-assisted translation. The tool was originally designed to capture data during post-editing of machine translation output by research subjects working remotely. TransCenter can record keystrokes and mouse clicks, and provides a 5-point Likert scale to allow participants to evaluate the adequacy of the proposed machine translation output. Time codes for each edit are captured by TransCenter as participants work, and then are displayed in reports generated by the researcher. In addition, the total amount of time spent by participants on each segment is recorded. Once a participant has completed the task, reports can be generated that show the progression of every edit, as well as overall statistics for each segment. Figure 1 illustrates the succession of events that comprise the participant's revision of a given segment. Each unique event is highlighted in the text with the corresponding time code appearing in the left-hand column. This type of report reconstitutes the typing and editing activity of each participant and shows the progression of the translator's work. Figure 2 reports segment-level data, summarizing the total number of edits, key presses, mouse clicks, the rating and time. These reports afford the researcher with global statistics for each segment.

Time	Sentence 9 Edits
Initial	The Fukushima disaster dealt a large blow to nuclear energy.
1389061464028	The Fukushima disaster dealt a large blow to nuclear energy.
1389061464028	The Fukushima disaster dealt a arge blow to nuclear energy.
1389061464371	The Fukushima disaster dealt a ablow to nuclear energy.
1389061464449	The Fukushima disaster dealt a dblow to nuclear energy.
1389061464652	The Fukushima disaster dealt a deblow to nuclear energy.
1389061464808	The Fukushima disaster dealt a devblow to nuclear energy.
1389061465135	The Fukushima disaster dealt a devablow to nuclear energy.
1389061465400	The Fukushima disaster dealt a devasblow to nuclear energy.
1389061465525	The Fukushima disaster dealt a devastblow to nuclear energy.
1389061465681	The Fukushima disaster dealt a devastablow to nuclear energy.
1389061465853	The Fukushima disaster dealt a devastatblow to nuclear energy.
1389061465993	The Fukushima disaster dealt a devastatiblow to nuclear energy.
1389061466118	The Fukushima disaster dealt a devastatinblow to nuclear energy.
1389061466196	The Fukushima disaster dealt a devastatingblow to nuclear energy.
1389061466399	The Fukushima disaster dealt a devastating blow to nuclear energy.
Final	The Fukushima disaster dealt a devastating blow to nuclear energy.

Figure 1: Example of TransCenter output: Keystroke logging and text edits (Participant 7)

ID	Source	Translation	Edited	Rating	Keypress	Mouseclick	Edits	Time
9	El accidente en Fukushima le dio un golpe mortal a la energía nuclar.	The Fukushima disaster dealt a large blow to nuclear energy.	The Fukushima disaster dealt a devastating blow to nuclear energy.	4	19	2	15	57293

Figure 2: Example of TransCenter Report: Segment statistics

Figures 1 and 2 present data generated by the same participant (Participant 7) working on the same sentence (segment 9). As described previously, each report provides a different perspective on the translation and editing process. While the work reported is the same, Figures 1 and 2 clearly depict the different perspective that each report supplies. The reporting capabilities of TransCenter made this tool the obvious choice for this experiment.

The original goal of TransCenter's creators to capture data on post-editing of machine translation is similar to the goal of this dissertation. Translation memory tools provide proposed translations much like machine translation systems do, and require the translator to edit the segment to render an acceptable target language version of the source text. TransCenter allows the researcher to specify the source and target language segments, so the texts that each participant worked with could easily simulate translation memory proposals. The built-in rating scale also allowed for participants to rate their perception of effort as they worked.

Before describing the experiment itself, it will be useful to briefly discuss the implementation of the tool. To use TransCenter for this experiment, Kent State University computing hardware and software were used to obtain consent and to host the web server. A one-question survey was created using Qualtrics and administered to potential subjects to obtain their consent to participate in the study. A copy of the consent text appears in Appendix A. After participants agreed to participate, their web browser was re-directed to a webpage hosted by a virtual machine running 64-bit Windows Server 2008 R2 Standard. The university IT infrastructure provided the requisite data security and storage needed to maintain confidentiality of the participants. Participants were assigned an anonymized participant number and unique log-in credentials to allow them to only complete the task once. Help documentation was provided by e-mail and was also available as a webpage before and during the task. After participants completed the task, the researcher generated the previously described reports, created a backup of each report, and stored the data in a password-protected archive.

Keystroke Logging

As mentioned in the previous subsections, keystroke logging is one of the data streams that TransCenter provides to the researcher. Regardless of whether keystroke logging is done on a local or virtual machine, the underlying assumptions as to its validity in process-oriented research remain the same. As Jakobsen (1999) indicates, there are two ways to gain insight into the translation process: direct observation and introspection. Direct observation using keystroke logging:

> …makes it possible to represent, and therefore to trace, the entire road along which a subject travelled before arriving at the final version of the target text. Instead of seeing only the final product […] we can observe all the underlying, preliminary layers of text and decision-making that contributed to the making of the final version. (Jakobsen 1999:12)

This complete record of the decision-making process helps elucidate the amount of effort exerted by translators during the translation process and the extent of the changes introduced by the participant. Without having a record of all the changes made, researchers risk misinterpreting any comparison made between the source and target text. In the context of translation memory tools, for example, an exact match may appear to have been left as is by the participant. Process data, however, might indicate that changes were initially introduced in the segment and subsequently removed to revert to the original segment. In this scenario, participants would have exerted effort to process and evaluate the proposed match, yet reviewing only the product data would suggest the contrary. Moreover, individual changes made within the text provide insight into how and why the effort is being applied. Likewise, product data might show changes at specific points within the text, while process data may suggest that additional areas initially required intervention.

Since time codes are recorded with each individual change, it is also possible to tabulate temporal information about each change. Temporal information provided by time-coded keystrokes is also indicative of effort. Extrapolating from Schilperoord (1996), Jakobsen (1999:14) remarks:

> The assumption underlying the recording of temporal information and the various ways of representing this information is (a) that there is a general correlation between time delay during text production and the cognitive processing that is involved and (b) that – more specifically – this correlation can be observed at different levels.

As noted, delays between keystrokes can correlate to cognitive processing, and by extension, effort. Likewise, the total time on task and time per segment provide information about the translator's progression through the text, and allow averages to be calculated of time per word or time per character. These measures are yet another way to visualize the distribution of time and effort throughout a text. Time delays or pauses are another type of temporal information provided by keystroke logging, which will be discussed in the following subsection.

Before addressing pause metrics in greater detail, brief mention should be made of the ecological validity provided by keystroke logging in experimental research. Capturing data related to the translation process in progress can be challenging, particularly when the research questions aim to understand the translator's behavior as it unfolds without changing the way in which the translator works. In an ideal scenario, keystroke logging is incorporated into a participant's typical translation workflow and remains undetected by the translator. The goal is to interfere as minimally as possible, while still having access to authentic process data. Unfortunately, research-oriented keyloggers of this type are not particularly plentiful. Moreover, these keyloggers when deployed in an online environment can be misidentified by security software as malware. Thus, programs such as TransLog and TransCenter require participants to work in a non-standard interface, but still allow them to complete the task without having to modify the way in which they work.

Pause Metrics

Pauses have been successfully used in research on speech production, writing, and revision as indicators of cognitive processing. Butterworth (1980:155–156) outlines the rationale for using pauses as evidence of cognitive activity: "the more the delays, the more cognitive operations are required by the output." This logic is grounded in decades of research on language production, and Butterworth understands the idea of a delay as "a period of silence, a pause" (ibid: 156). Translation studies scholars have also adopted the claim that pauses are evidence of cognitive operations (e.g., Krings 2001; Jakobsen 2002; Immonen 2006; O'Brien 2006b; Lacruz, Shreve, and Angelone 2012) and have incorporated pause metrics into translation process research.[11]

One particular way in which pause metrics have been successfully implemented in translation studies is as an indicator of cognitive effort in post-editing. For example, O'Brien (2006b:7) uses pause metrics as an indicator of "effort involved in post-editing MT output." O'Brien's study aims to identify correlations between source text machine translatability and post-editing effort, and triangulates pause data with Campbell's (2000) choice network analysis. The study is largely exploratory, yet her conclusions lend credence to the use of pauses as indicators of cognitive effort. O'Brien suggests that to obtain a more comprehensive view of the cognitive effort exerted by translators, pause metrics should be used in conjunction with other measures such as keystroke logging and choice network analysis.

Lacruz, Shreve, and Angelone (2012) build on O'Brien's research in an attempt to establish stronger correlations between pause data and cognitive effort. The authors contend that pauses in and of themselves do not sufficiently measure exerted cognitive effort, and therefore need to be examined in a different way. Lacruz, Shreve, and Angelone (2012:3) propose a new metric, the average pause ratio, which "is computed

[11] For a more comprehensive overview of pauses as an indicator of cognitive effort in process research of translation and post-editing, see O'Brien (2006).

for each segment as the average time per pause in the segment divided by the average time per word in the segment." This metric more closely aligns with the effort exerted by each pause and its relationship to the segment length. Their findings suggest this metric as measure that can approximate cognitive effort; however, the authors are quick to note that this metric needs to be tested in other situations before making generalizations about their findings.

Given the more nuanced approach to pauses provided by Lacruz, Shreve, and Angelone (2012), average pause ratios are adopted as one measure for studying the cognitive effort exerted by participants. To calculate pause ratios, however, the length of a pause must be established. Translation scholars have experimented with several pause lengths as potential thresholds for approximating cognitive processing. Immonen (2006) rightly asserts that written text production pauses necessarily differ from oral speech production. Butterworth (1980) notes that at the time of his writing on oral speech production that the minimum pause length would be 0.20 seconds to constitute a pause. Nevertheless, when using a keyboard to produce written text, this threshold more closely approximates the typing speed of many participants (Jakobsen 1998 in O'Brien 2006b). As a result, this dissertation adopts a one-second pause length in line with several other translation process researchers to allow for comparability of results.[12] By using this pause metric in conjunction with keystroke logging, this dissertation better approximates the

[12] Pause length tends to vary from study to study, but a number of researchers have used one second as a common threshold. Some more recent studies that adopt this duration are: Jakobsen 1998; Krings 2001; O'Brien 2006; and Lacruz, Shreve, and Angelone 2012.

participants' cognitive effort during translation and editing. Moreover, the average pause ratio as an indicator of cognitive effort can be tested.[13]

Segment Scoring

The previous two subsections on keystroke logging and pause metrics outline the rationale and methodology for measuring cognitive effort. These measures also allow the researcher to address the first research question posed in this dissertation: does the use of translation memory affect the cognitive effort of the translator during translation? The second question, though—does the translator's perception of the effort required during the task align with the actual effort exerted?—requires an additional data point. To ascertain what translators perceive about each segment, participants need to provide their opinion. The way in which the translator is asked to provide his information though requires careful consideration.

Think-aloud protocols (TAPs) and retrospective interviews are two methods that have been used in translation process research in an attempt to investigate cognitive behavior of translators.[14] Both methods rely on translators verbalizing the perception of their own behavior during the translation process; however, the verbalization occurs at different times. In TAPs, translators report their activity during the process itself, while

[13] Lacruz and Shreve (2014) have subsequently replicated the findings reported in Lacruz, Shreve, and Angelone (2012).
[14] An extensive overview of 108 works spanning two decades is provided in an annotated bibliography on TAPs in translation research by Jääskeläinen (2002).

retrospective interviews occur after the task has concluded. As noted previously, the aim of this study was to approximate, as closely as possible, a professional translation task and to not introduce additional tasks that are not typically required in professional translation tasks. Consequently, TAPs were not used to study the translator's behavior, as this additional task is not required of practicing translators.

Retrospective interviews were also not conducted in light of time considerations and a lack of compensation available to participants. These interviews would have required some kind of stimulus (e.g., interview questions or a screen recording) to elicit responses related to the task at hand. Moreover, no guarantee exists that the responses given would sufficiently elucidate the effort required of each segment. In addition, interviews would need to be conducted shortly after completing the task and would need to be scheduled by the researcher. Many of the participants completed the experiment between professional projects and were unsure of their availability. Consequently, to provide the greatest flexibility to participants and given the asynchronous nature of the study, retrospective interviews were not feasible and ultimately were not adopted.

To date, very few studies have been conducted concerning the perception of effort as it correlates with the effort exerted by the translator or editor. One exception to this dearth of research is Koponen (2012:181), which aims to investigate "the relationship between technical post-editing effort and cognitive post-editing effort by comparing cases where the edit distance and a manual score reflecting perceived effort differ." To elucidate perceptions of effort, participants were asked to rate segments based on the relative amount of effort required to edit them, using a five-point Likert scale. The researcher then compared the subsequent post-editing operations of participants to the scores. The researcher examined specific cases in which fewer post-editing operations were performed than originally anticipated by participants. Results from the study suggest that segment length may play a role in shaping the cognitive effort required to identify errors. In addition, segments that required syntactical reordering were perceived to be more effortful than segments that required only lexical or word-level changes.

The use of a numerical score to assess perceived effort is advantageous for a number of reasons. First, participants are restricted to a finite set of responses, ensuring that the researcher obtains the information required for analysis. Second, perceived effort is highly individualistic, and a numeric rating allows a participant to evaluate each segment individually.

Participant Selection

To minimize the number of potential participant variables that could confound the results and to create a relatively homogenous group, specific requirements were determined prior to conducting the study. Variables that were established include language direction, years of experience, and income derived from translation. An overview of the participants is provided below, followed by a description of the survey used to identify participants. The rationale for including each of the variables is laid out in the final subsection of this chapter.

Description of Participants

A total of nine translators participated in the study. The participants are Spanish-to-English translators with 4 to 7 years of experience, whose native language is English. Experience was defined as deriving a minimum of 30% of one's income from translation in a given year. The range of years is an arbitrary selection; however it is not randomly defined. Many of the previously-mentioned studies have focused on translators that are new to the field, or conversely, fall on the other end of the spectrum, having 10 or more years of experience. While both of these groups merit investigation in their own right, the idea behind this study is to constrain the amount of experience that each translator has in order to limit experience as a participant variable. Moreover, translators with between 4 and 7 years of experience should be familiar with computer-assisted translation tools, while not having so much experience that their approach to translation is completely automatized.

In addition, previous research has found a correlation between post-editing performance and years of experience. For example, Almeida and O'Brien (2010:7) specifically investigate this relationship and suggest that more experienced translators implement a "high number of preferential changes." Koponen (2012) also notes variability in the participants' approach to editing segments. Since this dissertation investigates cognitive effort when translating and editing proposed translations, large variations in editing behavior due to individual preferences could have skewed the results. To limit the discrepancy between participants, a three-year band of experience was established to mitigate for any potential correlation between years of experience and preferential changes.

Specifying the native language of the participants is in line with previous translation research (e.g., Almeida and O'Brien 2010; Carl et al. 2011) and aims to minimize variations arising from the participant's native language. Professional organizations (e.g., ATA) and standards (e.g., ASTM 2575) often suggest that translators should work into their native language, although some translators offer services in both directions. Such is particularly the case with languages of limited diffusion. The issue of directionality in translation is outside the scope of this dissertation; however, given that there are many translators working from Spanish into English, we chose to avoid any potential confounds by working with native English speakers.

Income derived from translation was the third variable that was controlled for in participant selection; participants in this study derive a minimum of 30% of their income from translation. This minimum threshold was determined to avoid selecting as participants translators who only occasionally provide language services. Participants were not compensated for their work, and were encouraged to approach the tasks as they normally would. Participants self-selected to take part in the study, and were aware that compensation would not be provided from the outset of the recruitment process.

The familiarity that participants have with a particular piece of translation memory software is not a confounding variable in the study, as the particular translation memory and keylogging software used in this study is TransCenter, developed by Denkowski and Lavie (2012). This server-based software is available for free but not widely used in commercial projects, and is ostensibly not a computer-assisted translation tool. The tool itself, as previously discussed, was originally designed to evaluate effort of post-editing machine translation output, and therefore is not a piece of software that any of the participants would have used. As a result, experience with one particular type of computer-assisted translation software did not provide any advantage to the participants. To mitigate this unfamiliarity with the software, instructions were provided via e-mail prior to the study. Participants were asked if they had any questions prior to taking part in the study. In addition, documentation was provided to the participants online prior to starting the study, and could be accessed at any time during the translation task. Translators were not provided any training and did not complete a practice session.

Pre-selection Survey

To identify participants who were suitable for this study, a pre-screening survey was used to select translators based on participant variables. This protocol allows for the establishment of a more homogenous participant group. The survey is imperative to ensure that confounding variables were not introduced that might have drawn attention away from the object of analysis. The pre-selection survey was conducted online to allow for translators throughout the U.S. to participate. The survey was created and distributed using the software package Qualtrics. The 13 questions that comprise the survey are presented in Appendix B. A link to the survey was circulated on several professional listservs, including that of Espalista, the American Translators Association Spanish

Language Division. Several professional translators who publish high-traffic blogs also posted the recruitment script (Appendix C) on their respective blogs. In addition, several professional translators and translation agencies agreed to circulate the call for participants to colleagues and to internal translation vendor lists. After completing the survey, participants who qualified for the experiment were contacted by the researcher to complete the study.

Experimental Task

Based on a review of the literature, keystroke logging, pause metrics, and segment scoring were the measures selected to test the research questions in this dissertation. A modest experimental study was conducted to expand on previous empirical research on TM use. TransCenter prompted participants to translate a text segment by segment and rate each segment’s difficulty. Participants completed this task over the Internet using TransCenter. Participants had unique login credentials that allowed them to access the task and complete the translation at their convenience. No time limit was imposed, although participants were asked not to consult any external references or resources.

The source text was the same for each participant, and was an excerpted newspaper article from a major Argentinean newspaper, *El Clarin*.[15] The article covered a visit by Japanese Prime Minister Shinzo Abe to Argentina in October 2013 and a press

[15] The entire article was 626 words, and the excerpt was 400 words, divided into 21 segments.

conference on the Fukushima nuclear meltdown. A copy of the source text is provided in Appendix D. The article was divided at the sentential level into 21 segments.

Each segment was presented in one of three conditions:

1. The segment did not have any proposed translation from a translation memory. Translators were required to translate the TL segment from scratch.
2. The segment was presented with a fuzzy match of 75–99%. Translators edited the presented fuzzy match to produce a TL segment that they deemed an acceptable rendition.
3. The segment was presented with a 100% match. Such segments were presumed to be acceptable renditions of the SL segment, and required the translator to verify that the translation was acceptable. Translators were still be given the option to edit the segment if they deemed it to be necessary, as would be the case in a professional project.

In all three conditions, translators used the TransCenter Likert rating scale from 1 to 5 to rank their perception of the relative difficulty of translating or editing each segment. A score of 1 meant that the translation provided was incomprehensible and needed to be completely retranslated, and a score of 5 meant that the proposed translation was perfect and did not require any edits.[16] In total, the presented text contained 21 segments, with 10 segments in each condition. In this way, translators worked with one

[16] The score values are as follows: 5 (Very good): Translation is perfect and does not require any edits; 4 (Usable): Translation contains a few errors, but editing is definitely easier than re-translation; 3 (Neutral): Translation contains more errors. It is difficult to say if it would be easier to post-edit or re-translate; 2 (Non-usable): Translation contains significant errors. It would be easier to re-translate the source sentence; 1 (Gibberish): Translation is incomprehensible.

text that contains segments that were to be translated without any aid from a TM, segments that have a TM solution that are acceptable, and segments that have a TM proposal that requires revision. This setup mimics a typical translation task that a professional translator regularly encounters, in which a portion of the text will require translation from scratch, while other segments have a proposed match that could be accepted as is, or that needs to be revised.

To avoid order effects and any confounds resulting from the stimulus, the three types of segments were interspersed throughout the text, so as to negate any judgment about the overall quality of the proposed translation. For example, the proposed segments in the third condition were not presented concurrently or clustered in one portion of the document, but rather were distributed throughout the text. This design replicates working conditions of practicing translators who occasionally encounter TL segments that require a significant amount of revision in order to create an acceptable translation. To allow individual segments to be compared across participants, three versions of the proposed TM matches were created. That is to say, for every source segment, there was an exact match, a fuzzy match, and a no match segment. From these segments, three different versions (Appendix E) of the stimulus were created. Appendix F shows the type of match presented to the participant in each of the three versions.[17]

During the task, the translator's keystrokes and mouse clicks were logged and the overall time per segment was recorded, as well as the time code associated with each

[17] As with the survey, participants were contacted via e-mail to invite them to participate. The recruitment script is included in Appendix G. Consent was obtained using a one-question Qualtrics survey question, and is included in this dissertation as Appendix H.

change. The time codes allow for the calculation of time delays and pauses between key presses. The rationale for recording pauses and overall time per segment has been discussed in the previous sections and is in line with the justification provided by scholars in research on cognitive effort. Moreover, using these same parameters will allow for comparability of results with previous studies to see whether support is lent to prior results.

Analysis

Given the large amount of data provided by TransCenter, the decision was made to use both inferential and descriptive statistics. Moreover, general observations of patterns and participant performance provide a panoramic view of the results. Statistical significance has been calculated between the three conditions (fuzzy and exact matches and novel translation) as well as any possible interaction. Likewise, pause patterns have been analyzed manually to determine if indicators of cognitive effort appear in different locations throughout the segments, and to see if effort is distributed differently in the three conditions. Average pause ratios have been calculated as an indicator of cognitive effort. Finally, the translators' rating of perceived difficulty has been analyzed to explore potential correlations between their perceptions of effort and their performance of the translation and editing task.

After all of the participants completed the task, data analysis began. The results of these analyses are outlined in Chapter 4.

CHAPTER 4

Results

Data Preparation

As discussed in the previous chapter, TransCenter was used to record all of the keystrokes, mouse clicks, and rating scores of the study participants. TransCenter generates several different reports for each participant. One provides an overall summary of the participant's behavior in the task, including the total number of mouse clicks, keystrokes, ratings, and time spent working in each segment. Another report presents the same information in aggregate for all of the participants. A third report, which is specific to each participant, provides an overview of all of the edits in each segment and visually represents these in tables with time codes associated with each change. Initially, the time elapsed between each time code was assumed to be the delay between each edit, and subsequently could be considered a pause; however, upon closer inspection, the visual representation of the segments was found to be misleading.[18] Consider the example shown in Figure 3, which shows Participant 4's edits in segment 2.

1388865939054	Elated at having achieved what he had come for, the representative claimed the crisis unleashed at the Fukushima nuclear power plant was absolutely under control.
1388867751533	Elated at having achieved what he had come for, the representative claimed the crisis unleashed at the Fukushima nuclear power plant was absolutely under control.

Figure 3: Participant 4, Segment 2 – Example of time codes

Here, the two edits appear one after the other in the report. The time elapsed between these two edits can be measured by subtracting the first time code from the second. The resulting difference is approximately 1800 seconds or 30 minutes (1812479 milliseconds). This difference is highly improbable, considering the fact that the participant spent approximately 43 minutes on the entire task. To better understand where this large time span originates, the researcher consulted the raw data stored by TransCenter. This data is saved in an SQL database, with each data type — events, ratings, counts, edits, and task statuses — being stored in a discrete table. The events table lists every change made by the participant during the task in sequential order. Upon closer inspection, this table showed that these two segments in Figure 3 were separated by more than 1900 events. In fact, the participant translated segments 3 through 21 after

[18] During the initial setup of the study, the researcher assumed that the TransCenter reports would provide all of the information needed to calculate pauses. This methodology is described in Chapter 3. After running the experiment, however, the researcher realized that the obtained data would need to be prepared before analyses could be conducted. As a result, the decision was made to include a data preparation section in Chapter 4, since the researcher realized that the adjustment was necessary after conducting the study.

the first edit shown in the figure before returning to the segment to make the second change. Consequently, a different approach than was initially planned was needed to determine the time delays between every edit, and would require the use of the raw data stored by TransCenter.

The events table, though, only includes the keystrokes and mouse clicks generated by each participant; the time codes for each rating do not appear in this table. These ratings are important, as are the times, because they better delimit each segment. In addition, the relative amount of time spent prior and subsequent to each participant providing a rating of each segment provides a more complete picture of the participant's cognitive processing. That is to say, the time the participant spends working within the segment is only part of the cognitive processing that occurs; the participant begins to pre-process each segment before translation or post-editing. To illustrate this point, consider the following example.[19] Participant 4 translates and edits segment 3 and provides a rating. Once the translator has scored segment 3, he begins to read and process segment 4 before introducing any changes.[20] Then, he makes all of the changes needed to complete the translation. Finally, the translator pauses to provide a rating. This time delay corresponding to this pause is assumed to be the time required to read and review the translation before providing a score. The post-processing time is also cognitive effort that the participant exerts and should be included in the overall pause time for the segment.

[19] This is a fictitious example, and serves merely to clarify how pre-processing and post-processing times were measured.

[20] The researcher recognizes that this upfront time may not account for a reading of the ST segment in its entirety and the proposed translation, as the translator may begin to translate after having only read part of the source text. No attempt is made to correlate the initial pause that occurs before any changes are made with reading time; instead, this processing time is included in the overall pause time for each segment.

To capture all of the processing effort, the researcher combined and collated the events and ratings tables in an Excel table to calculate all of the time delays between each event. To do so, the researcher first exported the events table and ratings table in CSV format. Then, the data was imported into Excel and sorted by participant. The events and ratings, along with the concomitant time codes, were placed in the same spreadsheet and sorted by time code. Once in chronological order, the difference was calculated between each event. Then, the researcher separated all of the time delays by segment. In this way, the time difference between each event was calculated, which rectified the 30 minute pause described previously and incorporated the pre- and post-processing time in the total time calculation for each segment.

As described in Chapter 3, pauses are considered to be time delays of one second or greater, and include the time that elapses before any editing within the segment has occurred. The pauses that occur after editing each segment has been completed but before the translator has rated said segment are also included, since the participant is reviewing the translation and thus still working with said segment. Time delays subsequent to the selection of an effort rating are deemed to be part of the next segment. Occasionally, participants adjusted a rating for a segment after having moved to another portion of the text. The time delays corresponding to these adjustments are included with the segment being rated, as the translator is assumed to have reviewed that segment again before making the adjustment. The inclusion of all of the intra-segment and inter-segment time delays provides a more complete picture of the effort being exerted for each segment.

General Observations

All of the participants successfully completed the task. None of the participants requested additional information or instructions, and none reported any problems during the task itself. The mean time for participants to complete the task was 48.73 minutes (SD 17.66), with times ranging from 26.9 minutes to 73.2 minutes. Table 1 reports some general information about the results of the translation task.

Table 1: General Overview of Experimental Task: Time (in minutes)

Mean	48.73
Median	43.20
Standard Deviation (SD)	17.66
Range	46.31
Minimum	26.88
Maximum	73.19

All of the participants were able to complete the task well within the allotted time of 90 minutes, although considerable variation is apparent across the 9 participants.

As noted above in the data preparation section, participants worked through the text sequentially, although the way in which they progressed differed. Several participants worked from start to finish, never returning to previous segments. These participants rated each segment as they translated, and moved through the file in strictly sequential fashion, segment by segment. Most participants, though, returned to a previous segment to edit the segment further or to change the rating. Occasionally, these regressions within the text were only separated by a few segments. Once participants had finished making any changes, they then resumed work in the remaining segments. The third group of participants completed the translation task and then revisited various segments throughout the text to introduce additional changes. Drawing on Waes' and Schellens' (2003) classification of writer profiles, one could argue that *non-stop writers* are comparable to the translators who moved from start to finish relatively quickly, and who did not introduce edits after the fact. *Fragmentary writers*, on the other hand, could be the participants that paused often and edited as they moved through the text. Further analysis is required to truly classify each participant; however, the participants' behavior clearly indicates differences in their writing and editing behaviors.

Exerted Effort by Match Type

To determine statistical significance of the effect that the match type — exact, fuzzy and new matches — has on the translator's cognitive effort, as measured by the average pause ratio, a one-way repeated measures ANOVA was used. Johnson (2008: 121–122) justifies this test for an experimental design similar to the one used in this dissertation, since the measures in each of the three conditions (fuzzy, exact, and novel matches) are all generated by the same participant. This test is somewhat more sensitive than a traditional ANOVA, because the individual's performance on the task is assumed to be more consistent across the three conditions. That is to say, individual differences in participants may result in larger variations in the obtained data, but comparing participants against themselves under the three conditions will provide a more accurate reflection of the treatment's effect on the dependent variables. Saldanha and O'Brien (2013:145) note, however, that tests like the ANOVA have "come under criticism in translation process research as such tests assume a high degree of control over all variables." In particular, Saldanha and O'Brien (2013) cite Balling's (2008) work, in which Balling makes the case for multiple regression designs.

Balling acknowledges the benefit of being able to determine statistical significance beyond individual differences by using regression models. Indeed, a lack of variable control can result in large individual differences between participants. Moreover, small sample sizes can emphasize these differences, making statistical significance more challenging to detect. Balling (2008:176) asserts that "factorial designs require strict control between groups of experimental items and therefore make more naturalistic, less experiment-like approaches difficult, if not impossible." In spite of these challenges, the researcher in this dissertation controlled for a number of participant variables to limit large variations in performance. Moreover, controlled stimuli were presented to each participant, which justifies the use of a factorial design.

The data were analyzed in a one-way, repeated measures ANOVA to compare the effect of match type on the average pause ratio; the three levels of the independent variable match type were new, fuzzy, and exact. Mauchly's test indicated that the assumption of sphericity[21] had been violated (χ^2 (2) = 20.822, $p < .001$), so the degrees of freedom were adjusted using the Greenhouse-Geisser estimate ($\varepsilon = .776$). There was a

---

[21] Sphericity is one of the assumptions that must be met when running a one-way repeated measures ANOVA (Maxwell and Delaney 2004).

significant effect of match type on average pause ratio, F(1.55, 96.185) = 4.556, p = .02. The means, sample sizes, and standard deviation of each match type are listed in Table 2. Large standard deviations for *Exact* and *Fuzzy* matches show sizeable overlap between these two match types. The average pause ratio of *New* matches, in contrast, has a much tighter distribution given the SD of .605.

Table 2: Descriptive Statistics – One-Way Repeated Measures ANOVA for Average Pause Ratio

Descriptive Statistics			
	Mean	SD	N
Exact	9.6647	18.67625	63
Fuzzy	6.9544	20.04157	63
New	1.2283	.60576	63

Pairwise comparisons showed a significant difference between *Exact* and *New* matches, while *Fuzzy* matches did not show statistically significant results. Estimated marginal means show significant overlap between *Exact* and *Fuzzy* matches. *New* matches have a much tighter grouping than *Exact* and *Fuzzy* match estimated means. The estimated marginal means are reported in Table 3 and the pairwise comparisons are presented in Table 4.[22] The estimated marginal means are presented in Figure 4.

[22] For reference, the categories indicated in the SPSS tables are as follows: 1 – Exact matches; 2 – Fuzzy matches; 3 – New matches.

Table 3: Estimated Marginal Means for Average Pause Ratio

Measure: Pause Ratio				
Category	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1. Exact	9.665	2.353	4.961	14.368
2. Fuzzy	6.954	2.525	1.907	12.002
3. New	1.228	.076	1.076	1.381

Table 4: Pairwise Comparisons for One-Way Repeated Measures ANOVA

Pairwise Comparisons						
Pause Ratio Measure						
Category	Category	Mean Difference	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	2.710	3.534	1.000	-5.986	11.407
	3	8.436*	2.349	.002	2.657	14.216
2	1	-2.710	3.534	1.00	-11.407	5.986
	3	5.726	2.534	.082	-.509	11.962
3	1	-8.436*	2.349	.002	-14.216	-2.657
	2	-5.726	2.534	.082	-11.962	.509

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
[b]. Adjustment for multiple comparisons: Bonferroni.

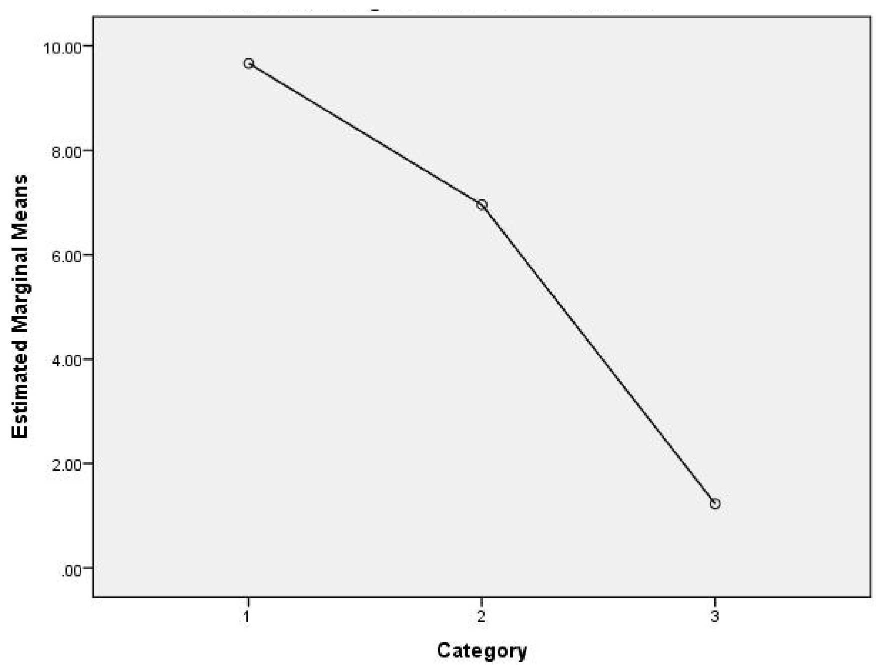

Figure 4: Estimated Marginal Means of Pause Ratios

In addition to the analysis of the average pause ratio, the data were also analyzed using a one-way, repeated measures ANOVA to compare the effect of match type on the various technical and temporal measures of effort. These measures include the number of keystrokes, the number of mouse clicks, the total amount of time spent in a pause, and the total amount of time spent per segment. A summary of these results is provided in Appendix I.

Perception of Effort vs. Keystrokes

Both Pearson and Spearman correlations were run for each of the participants to determine a potential correlation between the perceived effort (rating) of each segment and the keystrokes required for each segment. Five of the nine participants showed a significant correlation between the rating given to each segment and the number of keystrokes that were required to edit/translate the segment.[23] The individual values for each correlation appear in Table 5. Correlations that are significant at the .05 level are indicated with one asterisk (*) and those significant at the .01 level are indicated with two (**).

Table 5: Correlations by Participant – Rating and Keystrokes

Participant	**Pearson**	**Spearman**
1	-.120	-.079
2	-.746**	-.816**
3	-.195	-.282
4	-.584**	-.611**
5	-.528*	-.462*
6	.502*	.649**
7	-.814**	-.820**
8	-.081	-.305
9	-.126	-.325

Participants show strong significant correlations between their ratings and the number of keystrokes used in each segment. Most participants who showed a correlation demonstrated an inverse relationship between rating and keystrokes; that is to say, a higher rating (less effortful) required fewer keystrokes. Contrary to this pattern, though, is the correlation presented by Participant 6. A strong correlation between keystrokes and the rating was present; however, this correlation is in the opposite direction of all the

---

[23] If we consider the Spearman test, 5 participants also showed correlations that were significant at the 0.01 level (2-tailed). Pearson correlations are reported in the body of the text for comparability with the results of most other studies. Process-oriented research in translation tends to favor reporting Pearson correlation coefficients; however, Spearman's rank order test is more appropriate for rating scales that have discrete values. Any discrepancies when running rank-order will appear as footnotes.

other participants. This could be due to an inverted understanding of the Likert scale that was used; however, this data was not discarded from the analyses that were run as a whole, and did not change the levels of significance obtained. The participant-specific results do not show any change when these results are corrected.

In addition to participant-specific correlations, the same analyses were run based on the match type of each segment. Fuzzy matches showed a significant correlation between rating and keystrokes $r(63) = -.417$, $p < .01$, as did exact matches $r(63) = -.415$, $p < .01$. New matches did not show significant results.

Perception of Effort vs. Mouse Clicks

Of the nine participants, three showed a correlation between the ratings and the number of mouse clicks in each segment.[24] The specific values for each correlation appear in Table 6. Statistically significant correlations at the 0.05 level are marked with a single asterisk (*) and correlations that are significant at the 0.01 level are indicated with two asterisks (**).

Table 6: Correlations by Participant – Rating and Mouse Clicks

Participant	Pearson	Spearman
1	-.865**	-.862**
2	-.105	-.128
3	-.379	-.517*
4	-.279	-.246
5	-.454*	-.461*

[24] Four of nine participants show a correlation using Spearman's rank-order test.

6	.515*	.587**
7	-.028	-.053
8	-.212	-.244
9	-.307	-.415

As in the case of the correlations between rating and keystrokes, all but one of the correlations were inverse, with lower ratings (greater perceived effort) requiring more mouse clicks. Participant 6 again showed a positive correlation.

Once more, correlation statistics were calculated for each match type. Fuzzy matches showed a significant correlation between rating and mouse clicks $r(63) = -.309$, $p < .05$, as did exact matches $r(63) = -.513$, $p < .05$. New matches did not show significant results.

Perception of Effort vs. Time in Segment

Three of the nine participants showed a correlation between their ratings for each segment and the amount of time spent in the segment.[25] Table 7 presents the specific correlation values for each participant. Correlations that are significant at the .05 level are indicated with one asterisk (*) and those significant at the .01 level are indicated with two (**).

Table 7: Correlations by Participant – Rating and Time in Segment

Participant	**Pearson**	**Spearman**
1	-.150	-.461*

[25] Five participants showed this correlation when run using Spearman's rank-order test.

2	-.347	-.444*
3	-.267	-.444*
4	-.674**	-.628**
5	-.510*	-.405
6	.409	.428
7	-.669**	-.751**
8	-.391	-.385
9	-.203	-.285

As in the previous cases, an inverse correlation is observed, where higher ratings (less effortful) correlate with lower times in segments.

When correlations were calculated for the specific match levels, statistically significant results were again found in fuzzy and exact matches. Exact matches strongly correlated with segment time, $r(63) = -.455$, $p < .01$, while fuzzy matches correlated at the .05 level, $r(63) = -.296$, $p < .05$. New matches did not show a statistically significant correlation.

Perception of Effort vs. Exerted Effort

In this subsection, the exerted cognitive effort is compared to the translators' perception of effort. Exerted effort is measured by the average pause ratio, which was discussed in chapter 3. The perception of effort, as in previous subsections, continues to be assessed using the rating given to each segment. Correlations were found between the ratings of perceived effort and the actual exerted effort of three of the participants.[26] All of the correlations of the ratings and the average pause ratio are presented in Table 8.

When the correlations were calculated for match type, both exact and fuzzy matches again showed positive correlations. Positive correlations in this case suggest that higher ratings correlated with higher average pause ratios. Exact matches showed a correlation at the .05 level of $r(63) = .307$, $p < .001$, while fuzzy matches correlated at the .01 level, $r(63) = .413$, $p = .001$. New matches did not show a statistically significant correlation with average pause ratio.

Table 8: Correlations by Participant – Rating and Average Pause Ratio

Participant	**Pearson**	**Spearman**
1	.207	.056
2	.767**	.810**
3	.275	.187
4	.171	.443*
5	.327	.424
6	.110	-.468*
7	.559**	.935**
8	.405	.317
9	.499*	.437*

Additional Statistical Information

In addition to the correlation statistics reported thus far, three additional correlation tables are provided to show the values for each of the match types. Significant values are indicated in the same manner as those shown previously. Two asterisks (**)

[26] Five participants showed correlations between perception of effort and exerted effort when Spearman's rank-order test was used.

denote statistical significance at the .01 level, and a single asterisk (*) indicates significant correlations at the .05 level. Specifically, Table 9 shows the correlation values for exact matches, Table 10 shows the values for fuzzy matches, and Table 11 shows the values for new matches. Additionally, all of the individual statistics for participants are reported in Appendix J.

Table 9: Summary Correlation Table – Exact Matches

		Rating	Keystrokes	Mouseclicks	Pause Ratio	Pause Time	Segment Time
Rating	Correlation Coefficient	1.000	-.417**	-.513**	.307*	-.456**	-.455*
	Sig. (2-tailed)	.	.001	.014	.000	.000	
	N	63	63	63	63	63	63
Keystrokes	Correlation Coefficient	-.415**	1.000	.838**	-.170	.732**	.794**
	Sig. (2-tailed)	.001	.	.000	.182	.000	.000
	N	63	63	63	63	63	63
Mouseclicks	Correlation Coefficient	-.513*	.663**	1.000	-.184	.821**	.828**
	Sig. (2-tailed)	.000	.000	.	.148	.000	.000
	N	63	63	63	63	63	63
Pause Ratio	Correlation Coefficient	.307*	-.779**	-.184	1.000	-.175	-.247
	Sig. (2-tailed)	.014	.000	.148	.	.171	.051
	N	63	63	63	63	63	63
Pause Time	Correlation Coefficient	-.456**	.697**	.821**	-.175	1.000	.963**
	Sig. (2-tailed)	.000	.000	.000	.171	.	.000
	N	63	63	63	63	63	63
Segment Time	Correlation Coefficient	-.455*	.766**	.828**	-.247	.963**	1.000
	Sig. (2-tailed)	.000	.000	.000	.051	.000	.
	N	63	63	63	63	63	63

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. Match = EXACT

Table 10: Summary Correlation Table – Fuzzy Matches

		Rating	Keystrokes	Mouseclicks	Pause Ratio	Pause Time	Segment Time
Rating	Correlation Coefficient	1.000	-.417**	-.309*	.413**	-.339**	-.296*
	Sig. (2-tailed)	.	.001	.014	.001	.007	.019
	N	63	63	63	63	63	63
Keystrokes	Correlation Coefficient	-.417**	1.000	.663**	-.779**	.697**	.766**
	Sig. (2-tailed)	.001	.	.000	.000	.000	.000
	N	63	63	63	63	63	63
Mouseclicks	Correlation Coefficient	-.309*	.663**	1.000	-.638**	.519**	.591**
	Sig. (2-tailed)	.014	.000	.	.000	.000	.000
	N	63	63	63	63	63	63
Pause Ratio	Correlation Coefficient	.413**	-.779**	-.638**	1.000	-.566**	-.651**
	Sig. (2-tailed)	.001	.000	.000	.	.000	.000
	N	63	63	63	63	63	63
Pause Time	Correlation Coefficient	-.339**	.697**	.519**	-.566**	1.000	.963**
	Sig. (2-tailed)	.007	.000	.000	.000	.	.000
	N	63	63	63	63	63	63
Segment Time	Correlation Coefficient	-.296*	.766**	.591**	-.651**	.963**	1.000
	Sig. (2-tailed)	.019	.000	.000	.000	.000	.
	N	63	63	63	63	63	63

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. Match = FUZZY

Table 11: Summary Correlation Table – New Matches

		Rating	Keystrokes	Mouseclicks	Pause Ratio	Pause Time	Segment Time
Rating	Correlation Coefficient	1.000	.044	.014	.026	.188	.123
	Sig. (2-tailed)	.	.732	.916	.842	.139	.336
	N	63	63	63	63	63	63
Keystrokes	Correlation Coefficient	.044	1.000	.376**	-.277*	.624**	.747**
	Sig. (2-tailed)	.732	.	.002	.028	.000	.000
	N	63	63	63	63	63	63
Mouseclicks	Correlation Coefficient	.014	.376**	1.000	.108	.441**	.320*
	Sig. (2-tailed)	.916	.002	.	.398	.000	.011
	N	63	63	63	63	63	63
Pause Ratio	Correlation Coefficient	.026	-.277*	.108	1.000	-.051	-.226
	Sig. (2-tailed)	.842	.028	.398	.	.690	.075
	N	63	63	63	63	63	63
Pause Time	Correlation Coefficient	.188	.624**	.441**	-.051	1.000	.903**
	Sig. (2-tailed)	.139	.000	.000	.690	.	.000
	N	63	63	63	63	63	63
Segment Time	Correlation Coefficient	.123	.747**	.320*	-.226	.903**	1.000
	Sig. (2-tailed)	.336	.000	.011	.075	.000	.
	N	63	63	63	63	63	63

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. Match = NEW

CHAPTER 5

Discussion of the Results

The results presented in the previous chapter provide substantial insight into the behavior of translators working with computer-assisted translation technology. The data can be analyzed from a number of different angles; however our focus will remain on the two research questions posed at the outset:

(1) whether the relative matching statistics of the translation proposed by the translation memory tool affect the translator's cognitive effort, and;

(2) whether the participant's perceived effort aligns with the amount of effort that the participant actually exerted.

Each question will be addressed in turn, with general observations being provided for each, followed by a discussion of the reported inferential statistics.

Effect of Match Levels on Cognitive Effort

General Observations and Initial Pauses

To frame the discussion of translator behavior, this dissertation adopts Angelone's (2010) tripartite behavioral model of uncertainty management in translation. The model has been used to describe the exertion of cognitive behavior in other studies (e.g., Lacruz, Shreve, and Angelone 2012) and will facilitate comparison of the results of this study with the findings of other studies.

Angelone's (2010) model describes three steps that a translator takes as he or she solves problems throughout the text. The first step, *problem recognition*, as its name implies, requires the translator to identify areas of difficulty in the text. Behaviors that indicate problem recognition can include keyboarding pauses and mouse repositioning.[27] Following problem recognition is the *solution proposal* step, which "consists of strategy planning and/or application, with the immediate objective of generating and, as Tirkonnen-Condit has suggested, 'trying-out' potential solutions for the encountered problem." (Angelone 2010:20) Solution proposal may manifest as reading potential options that are available, typing multiple solutions, or verbalizing potential options. The final step, *solution evaluation*, requires the translator to assess the generated text. Angelone also asserts that the translator may make this assessment while writing the

[27] Angelone (2010:20) notes additional behaviors that indicate problem recognition, including indecisiveness in think-aloud protocols, and eye-tracking data such as fixation points and eye regressions. Keyboarding data such as pauses and mouse repositioning are most pertinent to this dissertation.

target text. These three steps, termed “triadic bundles” by the author, individually elucidate the type of effort that a participant exerts.

With this model in mind, general comments can be made about the effort exerted by participants: cognitive effort is required in every segment, regardless of match type. Regardless of whether edits were made, every segment in the text required the translator to evaluate the segment to determine whether the proposed translation was acceptable. One manifestation of this cross-language evaluation is a pause occurring at the beginning of a segment. In all three conditions, every participant made an extended pause at the outset of each segment. This behavior mimics participant behavior observed by Lacruz, Shreve, and Angelone (2012: n.p.), who “frequently observed clusters of long pauses during the reflective stages of reading, problem recognition, and solution proposal, stages that place high cognitive demand on the post-editor.” These researchers focused on cognitive effort during post-editing of machine translation, although as argued in chapter 2, the task is comparable.

Again, Angelone’s (2010) terms help describe the translator’s behavior at the outset of segments. The first two steps — problem recognition and solution proposal — in particular explain an initial pause. When editing, translators necessarily read the source language segment and the corresponding translation. Then, translators identify differences between the source segment and its proposed translation, if any. Potential edits can then be made and evaluated. This process can occur at any point in the segment, although there is inevitably an initial pause as the editor begins to work. This cross-language evaluation is characteristic of the editing task.

Editing, though, is not the only task that a translator performs when working with computer-assisted translation tools, and specifically, translation memory tools. Instead, translators are also presented with segments that require translation. Consider the three types of segments that were presented to translators in this dissertation's experimental task. Fuzzy and exact matches require the translator to read the source segment as well as the target. These match types ostensibly require editing, and as noted previously, the observed results of these match types can be can be compared to the results presented in Lacruz, Shreve, and Angelone's (2012) work. In the absence of a proposed match, there is no target language segment to investigate. Instead, translators are required to provide a translation and then review it. Cross-language evaluation still occurs in this scenario, but at a different point than during editing. The extended initial pause observed in this dissertation's experimental task, then, can be attributed to two different, but related, cognitive processes.

When translators are working with fuzzy and exact matches, cross-language evaluation, problem recognition and solution proposal are assumed to be the cognitive processes in play. The problems being identified by the translator here are deviations between the source segment and the proposed translation. These deviations could stem from problems with terminological choices, syntax, grammar, punctuation, or cohesive devices. The solutions that are proposed and evaluated are those necessary to modify the target text.

Initial pauses when participants are translating no-match segments, in turn, are assumed to be the result of reading and comprehension. The translator is pre-processing the segment, presumably reading at least part, if not all, of the segment. The translator then renders the source text in the target language. When translating no-match segments, problem recognition is associated primarily with translation and secondarily with editing during the self-revision stage.

The purpose of this discussion of these initial pauses ultimately is to show that cognitive effort is required for every segment. Translation memory tools do not obviate the need to exert effort, although they alter the overall amount effort required and change how the translator works. Ideally, translators would not identify problems in exact matches, and would not modify these segments. Rather, translators would only identify problems in the fuzzy match translations and would then exert effort in the solution proposal and solution evaluation stages.

Editing Behavior

In this study problem recognition is not confined to fuzzy matches; instead, every participant edited both fuzzy and exact matches. The type of edits being introduced fall outside the scope of this dissertation, but further linguistic analysis of the complete editing events may prove beneficial to our understanding of what triggers revision in translation.[28] Moreover, a replication of the Lacruz, Shreve, and Angelone (2012) study

[28] Complete editing events, as defined by Lacruz, Shreve, and Angelone (2012:2) are a "[collection] of individual editing actions [that] can be considered to naturally form part of the same overall action." These editing events help distinguish between edits that may not be cognitively demanding but require a greater

would be fruitful, with particular emphasis being placed on the relationship between complete editing events and cognitive demand. In doing so, the overall demand of specific segments might be better correlated with the effort exerted by each participant.

In general, participants exhibited different writing and revision behaviors. These behaviors can be better described using a typology developed by Waes and Schellens (2003). The typology relies on revision behavior and pause patterns to describe writer profiles. Five different writer profiles are identified: the initial planner, the fragmentary Stage I writer, the Stage II writer, the non-stop writer, and the average writer.[29] The first profile, the *initial planner* is described as making "relatively few revisions" and making pauses of significant longer average duration than those of other types of writers (845). *Fragmentary Stage I writers* are those participants who revise as they complete their initial draft of a text. These writers revise more during the initial drafting phase and make relatively fewer changes once the draft is complete. These participants also show "a high degree of recursion" (ibid) in their writing process. *Stage II writers* show the opposite behavior, with the majority of their revisions occurring once the draft is complete. Pauses are infrequent, but are longer than other writer profiles. *Non-stop writers* revise less than any other profile and pause less often. These participants also make fewer revisions once the draft is completed. *Average writers* tend to be writers who do not neatly fall into any of the other writer profiles.

---

amount of technical effort to implement, and edits that are cognitively taxing but may require relatively little technical effort.

[29] Waes and Schellens (2003: 837) refer to Stage I as the time "from initiation to the completion of the first draft," while Stage II falls between "the completion of the first draft to the completion of the final version."

While these profiles are not specific to the translation process, they are useful in describing the type of behavior exhibited by participants who took part in this dissertation's experimental task. Some participants, for example, translated the text from start to finish and revised each segment before moving onto the next. Such participants moved between each segment sequentially, never returning to previously translated work. These translators rated each segment in turn as they completed them. In Waes and Schellens' terms, these translators might be considered *initial planners* or *fragmentary stage I writers*. Other participants skipped around the translation, and did not move in chronological order. These participants revisited previously translated segments as they moved through the text, and occasionally changed their rating of some segments. They also edited previous segments and appeared to review their work several times. A classification of *fragmentary stage I writer* or *stage II writer* might best describe these participants. Still others seemed to progress chronologically in chunks of several segments at a time. These participants worked on three or four segments in a row, and stayed within this group of segments until all of the segments had been translated, edited, and rated. Translators who exhibited this type of behavior did not move to the next group of segments before they were satisfied with those in which they were working. These translators might be considered *fragmentary stage I writers* or *non-stop writers*.

As can be seen, each of these behaviors potentially aligns with Waes and Schellens' (2003) writer profiles. The profiles that Waes and Schellens propose, though, are largely dependent on writing tools. Their study focused specifically on the difference between writing using a word processor and writing using paper and pen. Curiously, the constraints in this study were the same for all participants, but the writers still exhibit unique editing behaviors as they progress through the text. Further research into translator behavior, and analyzing the process data presented in this dissertation in light of specific translator profiles might highlight even stronger correlations between perceived effort and technical, temporal, and cognitive effort. Moreover, this type of work could differentiate how cognitive effort is exerted and how cognitive resources are allocated during the translation and editing tasks. In this dissertation, writing profiles were not established for each participant at any point, so a more detailed analysis of their behavior cannot be determined at this time. Nevertheless, the participants' editing and translation behavior suggests that translators do not all approach the task the same way and the preference for a particular editing behavior could ultimately affect the way cognitive effort is exerted.

In short, these general comments on participant editing behavior and initial pauses contextualize the results obtained during the experimental task. Moreover, greater differentiation of participant variables could provide additional insight into the results reported in chapter 4. That said, the results that are discussed here focus squarely on the statistical results as they relate to the two research questions. First, the match levels' effect on cognitive effort will be reviewed, and then the participants' perception of effort will be discussed.

Exerted Effort by Match Type

As described in chapters 3 and 4, the average pause ratio metric (Lacruz, Shreve, and Angelone 2012) was used as a measure of cognitive effort. Lower pause ratios are considered to be more cognitively demanding, while higher ratios are deemed to be less demanding. The statistically significant results reported in chapter 4 from the one-way repeated measures ANOVA contradict the researcher's initial hypothesis. The researcher expected fuzzy matches to require the greatest amount of cognitive effort, as measured by the average pause ratio. Exact matches were hypothesized to require the least amount of effort, with the amount of effort required by new translations falling between the two. Numerically, the researcher hypothesized the lowest ratio would be measured in fuzzy matches, while the highest would be measured exact matches. Instead, the participants' average pause ratio (1.22) was found to be lowest when working on no-match segments. The highest ratio (9.66) was calculated in exact matches, which aligns with the original hypothesis. The fuzzy match pause ratio of 6.95 fell between new and exact matches. A summary of the hypothesis and the results is presented in Table 12.

Table 12: Hypothesis and Results - Exerted Effort

Amount of Effort	Hypothesis	Results (Average Pause Ratio)
Least	Exact	Exact (9.66)
	New	Fuzzy (6.95)
Greatest	Fuzzy	New (1.22)

Consequently, exact matches were found to be the least cognitively demanding, while new matches were the most cognitively demanding. The mean average pause ratio for fuzzy matches fell between the two.

These findings support those reported by O'Brien (2006), who used the physiological measure of pupil dilation to determine the amount of cognitive effort exerted when presented with four types of matches. In that study, the stimuli were new, fuzzy, and exact matches, as well as machine translation output. O'Brien reports that exact matches required the least amount of effort, while new matches required the greatest. Falling between exact and new matches were the fuzzy and MT matches. O'Brien's study does not report statistical significance between the measures; however, this dissertation's findings support O'Brien's results.

Closer examination of the differences between match types presented in chapter 4 reveals that statistical significance was only found between new and exact matches. Fuzzy matches approached statistical significance when compared to new matches ($p = .08$), and were not statistically significantly different from exact matches ($p = 1.0$). This very high *p*-value in the comparison of fuzzy and exact match means draws attention to the fuzzy match type and raises several issues. The first question posed is whether the input was flawed in some way such that significant differences were not observed in the average pause ratios (APRs) of participants working on fuzzy matches compared to APRs of participants working on either exact matches or no-match segments. This possibility seems implausible, since each fuzzy match required some sort of intervention on the part of the translator, while exact matches did not. Moreover, translators were instructed at the outset of the task that they could accept any proposed translation as is or edit it as they wished. The assumption is that translations that do not require edits would not be changed.

If the input was not flawed, then, a second question arises: what could explain the large difference in *p*-values between the conditions? Participant behavior is one possibility that warrants discussion. As described above, translators need to edit fuzzy matches but do not ostensibly need to change exact matches. Nevertheless, participants did edit exact matches. This behavior was exhibited by all participants; moreover, each participant edited multiple exact matches. Consequently, the calculated average pause ratios of exact and fuzzy matches are quite similar. This possibility of translator behavior explaining the large difference in *p*-values is supported by the considerable overlap in the distribution of *Fuzzy* and *Exact* matches, as well as the estimated marginal means. In the experiment, the mean average pause ratio of *Exact* matches was 9.66 (SD = 18.67), with *Fuzzy* matches having an average pause ratio of 6.95 (SD = 20.04). The estimated marginal means showed smaller standard deviations (2.353 and 2.525, respectively), but still show considerable overlap. The averages and estimated marginal means are shown in Figure 5.

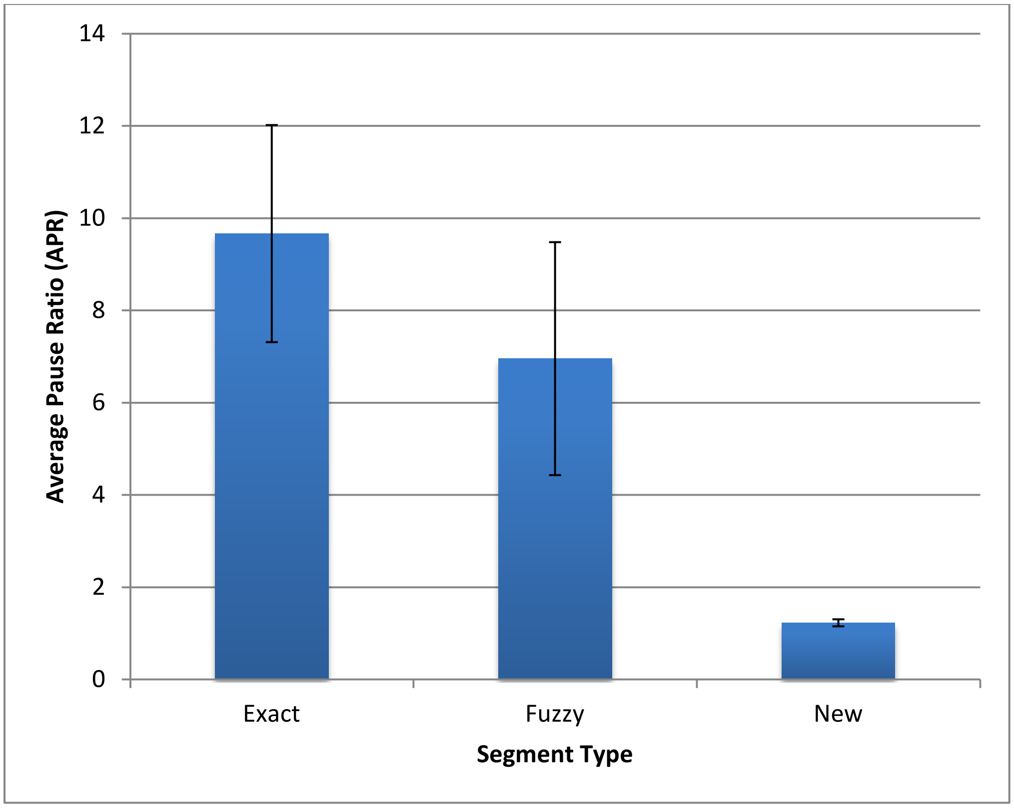

Figure 5: Estimated Means of Average Pause Ratio with Standard Error

The data presented in Figure 5 clearly indicate the effort exerted by participants when translating a segment from scratch is markedly different from the effort exerted when participants are presented with an exact or fuzzy match. These results support the argument made in this dissertation that the use of translation memory affects the cognitive effort exerted by translators. More specifically, the results indicate that participants exhibit a substantial decrease in the amount of effort exerted when editing a proposed translation compared to translation of a segment from scratch.

The similarity in the effort exerted by participants when working on exact and fuzzy matches merits further discussion. When considered in the aggregate, translator behavior seems to suggest that the overall approach to evaluating a proposed translation match is similar in light of the amount of edits introduced in both fuzzy and exact matches, and that translators do not necessarily differentiate between the two match types. Indeed, this particular experiment did not alert translators to the status of each match type during the task, which the translators may have taken to mean that all translations required editing.[30] Translators were allowed to accept translations as they were, though, and many did accept at least one segment without introducing any changes. As a result, translators were aware that proposed translations could be accepted as the final translation. Nevertheless, translators chose to treat exact matches as if they were fuzzy matches, adapting acceptable translations presumably to reflect the way in which they would have translated the segment. Consequently, translators did not show a significant decrease in effort between fuzzy and exact matches. Instead, translators evaluated all of the matches and changed the majority of the segments.

This type of behavior seems to constitute over-editing. Editing these matches is cognitively more demanding that translating a segment from scratch. Curiously, though, none of the participants deleted an entire proposed translation and started over. Instead, when participants revised a proposed translation, they continued to exert effort trying to edit the proposal to render it satisfactorily. Although the data do not support any firm conclusions, one possible hypothesis that explains participants' behavior of not abandoning the segment and continuing to edit the target language proposal would be the

[30] Several translation memory tools (e.g., SDL Trados Studio, memoQ, and Across Language Server) indicate the percent match level for each segment. Translators can still edit exact matches using these tools (unless the segments are locked by the project manager).

translator's supposition that translation memory matches are provided in support of the translation task. Many translators often anecdotally describe translation projects in which translation memories were faulty or impeded the translation process, and ultimately distrusted the source of the translations stored in the TM. In this dissertation, the participants seem to accept the proposed translations as an aid to their work rather than suggestions to be discredited or discarded. Further work on the translator's behavior when using these matches is needed to understand the threshold below which translation memory matches cease to be useful.

The aggregate statistics support the notion that exact matches are often treated as fuzzy matches, and that translators often over-edit exact matches; however, the results could also be influenced by order effects.[31] In the context of this study, order effects might be the result of the position of a particular match type in relation to others. For example, a fuzzy match that occurs after an exact match might be influenced by the participant's interaction with the preceding segment. Typically, experimental research designs aim to counterbalance stimuli with the goal of mitigating order effects. This type of counterbalancing, however, presupposes that each item in a study can be presented independently of the other items; items do not necessarily need to be shown in a specific order. This independence is not possible in this dissertation's experimental task since the text as a whole is being presented sequentially, on a segment-by-segment basis from start to finish. The experimental design of the dissertation did intersperse the different match

[31] Order effects are "the influence on a particular trial that arises from its position in a sequence of trials" (Heiman 2001:757).

types throughout the text, and ensured that the matches were not systematically presented in a given sequence or associated with specific segment. Both of these experimental design decisions were made to avoid large order effects resulting from the segment's overall position. As a result, the three types of segments can be presented to the translator in any possible combination. Even so, order effects are a real consideration that should be examined.

In the translation task, there are instances in which fuzzy matches follow an exact match. In such cases, one would expect participants to leave the exact match as is and then edit the subsequent fuzzy match. Some participants did exhibit this type of behavior. Others, however, did not. Instead, some participants first behaved as expected and accepted the exact match as it appeared. Then, participants accepted the following fuzzy match as well, but did not make required changes to render an adequate translation. The missing edits were typically minor, such as a punctuation mark or a subordinate conjunction. In a few cases, though, the translators failed to include information from the source segment. For example, one fuzzy match segment and its exact match counterpart appeared as follows[32]:

Table 13: Example of Fuzzy Match Segment

Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar.	Since then, there have been ~~small~~ leaks of contaminated water that have ~~wound up~~ ended in the ~~ocean~~ sea.

[32] The target language strikethroughs indicate the text that was removed from the exact match segment, while the underlined text shows text that was added. The researcher introduced one erroneous change and two preferential changes into the exact match to create the fuzzy match segment. The erroneous change is ostensibly a mandatory revision, while the others are not.

Table 14: Example of Exact Match Segment

Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar.	Since then, there have been small leaks of contaminated water that have wound up in the ocean.

The proposed translation does not include the adjective “small” and also has two other differences from the proposed exact match. The translator should have included the adjective in the translation, and perhaps made some of the other changes in the absence of the proposed exact match. These latter changes might have been considered optional, but the modifier should have been included. Nevertheless, when the translator, Participant 9, was presented with this segment, no changes were made. Instead, there seemed to be some carryover from the exact match that appeared just before the segment. Therefore, it is possible that the results have been skewed somewhat as a result of order effects that occur in the text.

Additional scrutiny in future studies is required to determine the impact that the order in which the fuzzy and exact match segments are presented has on the translator’s behavior. Experimental protocols to specifically test these order effects would need to be developed for this type of study, since the present study attempted in part to counterbalance these effects.

Perception of Effort vs. Keystrokes

The correlation statistics that were calculated for this subsection and those that follow address the second research question as to whether the effort exerted by participants correlates with perceived effort. As described in Chapter 3, perceived effort is measured by the 5-point Likert scale rating that participants provided for every segment. Effort, then, is the second measure against which the correlations have been calculated. Keystrokes are a measure of technical effort, and five participants showed strong correlations between perceived effort and the number of keystrokes required to edit or translate a given segment.

This correlation reveals that participants conflate technical effort with exerted effort. As discussed previously, technical effort is the physical manipulation of hardware to introduce changes and does not necessarily coincide with cognitive effort. Changes that require relatively few keystrokes may be considerably more demanding of cognitive resources than edits that require a greater number of keystrokes. In this sense, the translators that show a strong correlation between technical effort and perceived effort are not necessarily assessing the type of edits being introduced. Effort to these participants seems to be understood as a mechanical exercise, rather than a problem-solving one. If we consider this correlation in Angelone’s (2010) terms of uncertainty management, these translators perceive effort at the solution proposal and solution evaluation stages, rather than at the initial problem recognition stage.

The fact that some, but not all, of the participants conflate technical effort with exerted effort also suggests that translators are not always fully aware of their behavior. Consequently, the lack of behavioral awareness supports the need for metrics that take into account the type of edits being introduced, as well as segment length and word length, rather than using keystrokes as the sole measure of effort. Previous research that has used keystrokes may need to be reexamined using newer metrics, such as the average pause ratio, to validate previous findings.

The significant correlations shown by match type — fuzzy, exact, and no match — partially replicate the findings reported concerning exerted effort. On the whole, participants conflated perceived effort with the number of keystrokes when working with both fuzzy matches and exact matches. No correlation between perceived effort and keystrokes was observed when participants were working on no-match segments. Again, we see similarities between the perception of effort and the fuzzy and exact match types. Translators perceived little difference in the technical effort required by fuzzy matches versus exact matches, while the amount of technical effort does not correlate with the perception of the overall effort required by no-match segments. This finding is somewhat expected, since new translations will necessarily require the use of the keyboard, while the required editing could be easily conflated with keystrokes.

Perception of Effort vs. Mouseclicks

Just as some participants demonstrated strong correlations between perceived effort and keystrokes, four participants showed strong correlations between perceived effort and the number of mouse clicks. This correlation once again reveals a conflation of perceived effort with technical effort. Mouse clicks though, are a different behavior than keystrokes, and are not necessarily required when translating. This type of technical effort is arguably a more conscious action, since translators are not required to use the mouse to input translations. In addition, the relationship between mouse clicks and the perception of effort might reflect participants' navigational preferences. If participants prefer to navigate using a mouse, they may not consciously perceive mouse usage (and the concomitant technical effort) to be effortful. Conversely, participants who do not typically reposition their cursor using the mouse might well perceive mouse clicks to be more effortful. This study did not control for navigational preference as a variable. Thus, no conclusions can be drawn about the relationship between navigational preference and the perception of effort.

Once again, correlations were determined to be significant between mouse clicks in fuzzy and exact matches and perceived effort. Significant correlations were not observed between the number of mouse clicks and the perceived effort of translating no-match segments. The similarities between the type of effort that exact and fuzzy matches are perceived to require by participants emerge in these correlations as well, since the participants strongly correlated technical effort (as measured by mouse clicks) with the perceived amount of required effort. As with the keystroke correlations, the observation of correlations with both fuzzy and new matches, but not with new matches (no-match segments), is not unexpected, since new translations would not necessitate the use of a mouse.

These findings are also in line with those of Koponen (2012) who studied the relationship between the perceived effort of post-editing and the actual edits made by participants. Koponen's findings suggest that more technical and cognitive effort is required when editors are required to reorder segments. Participants who perceive mouse clicks as effortful may be exhibiting similar behavior, since mouse clicks may be the result of the act of re-ordering a segment. The data provided by TransCenter for this dissertation's study does not allow the researcher to ascertain if this is the case; however it is a plausible assumption that participants using the mouse were either repositioning the cursor or re-ordering the segment. In this light, Koponen's findings are supported by the correlations found between perceived effort and exerted technical effort in this dissertation.

Perception of Effort vs. Time per Segment

Unlike keystrokes and mouse clicks, the time spent by a participant in each segment is a temporal measure. Significant correlations between segment time and a participant's rating would suggest a conflation of temporal effort with cognitive effort. Again, Krings' differentiation between temporal, technical, and cognitive effort can be useful in understanding these results. Just as technical effort is one way to measure the underlying construct of cognitive effort, so too is the overall time spent completing the task. As a result, participants may liken temporal effort to the underlying cognitive effort required to translate or edit a given segment.

It should be noted that the correlation between time and effort is not necessarily a strict one-to-one relationship; more time does not necessarily mean more effort. For example, some edits may require more time simply because more keystrokes or mouse clicks are required, whereas little effort is exerted in the triadic bundle of problem recognition, solution proposal and solution evaluation. Longer segment times may also indicate a greater number of pauses, which is another potential measure of cognitive effort. Initial pauses and final pauses were both included in the calculation of overall time spent per segment. These pauses were included in an attempt to obtain a more complete picture of the temporal effort required of the translator. Moreover, including these initial and final time spans in the overall segment time most likely aligns with how participants perceive the time spent per segment. In Angelone's (2010) terms, participants are aware of the initial problem recognition phase and actively evaluate the proposed solutions. Consequently, participants will perceive the effort exerted for each segment not only as temporal and technical measures within the segment, but also as the entire time required to edit or translate.

Five participants show correlations between the measure of time per segment and perceived effort. Each of these participants also showed a correlation with a single measure of technical effort, whether keystrokes or mouse clicks, but never both. These participants are assumed to be aware of the technical effort required to translate or edit each segment and the amount of time needed to implement these changes. Temporal and technical effort is often intertwined, and these participants seem to demonstrate that the two types of effort are in fact interrelated.

When considering the correlations that occur when segments are compared by match type, the perception of effort and time per segment correlate most strongly in exact matches. This finding is indicative of the solution evaluation phase. Participants are acutely aware of the time needed to evaluate each segment and relate this evaluation phase to exerted effort. Fuzzy matches also show a correlation, but the perceived relationship is not as strong. In fuzzy matches, it can be assumed that the perceived effort derives from the triadic bundle. Again, the effort required to work through the problem recognition, solution proposal, and solution evaluation steps is comparable to the effort to perform these same steps when working with exact matches. Participants seem to differentiate, though, albeit slightly, between the effort required to only evaluate the exact and fuzzy matches. Participants, however, did not correlate the time needed to translate a given segment.

Perception of Effort vs. Exerted Effort

Having looked at how translators perceive technical and temporal effort, we now examine the relationship between what translators perceive and how much effort is actually exerted. Since temporal and technical measures are indicators of the underlying cognitive effort construct, the previous correlations provide insight into translators' perceptions of effort and their behavior. To investigate this relationship, the average pause ratio for each segment was calculated and then compared to the rating provided by each participant. Of the nine participants, five showed a positive correlation between the two measures. That is to say, segments that were rated highly also had the highest pause ratios. Keeping in mind that higher average pause ratios indicate less effort, these five participants rightly correlated the effort required to edit or translate the segment with the amount of effort that they exerted. The remaining four did not show this result.

Of the five who correlated perceived effort with exerted effort, four also correlated perceived effort with one or the other measure of technical effort and with the time spent in the segment. These participants seem to exhibit a metacognitive awareness of their translation behavior, and seem to consciously assess the type of effort required to edit or translate a given segment. The lone participant who showed a correlation between perceived and exerted effort, but not with technical or temporal measures, may perceive effort differently. This participant perhaps considers the level of intervention required to edit or translate the segment, and considers the technical or temporal effort a necessary component of the translation task. Data is not available to determine the exact cause of this variation; however, it may serve as the foundation for future work on the relationship between the three categories of effort in Krings' tripartite model.

As for the remaining participants whose data did not demonstrate a relationship between exerted and perceived effort, one possibility is that the translation task was not considered to be particularly effortful. Consequently their ratings cluster around several values and therefore do not show any relationship. Future studies will attempt to mitigate for this potential confound by increasing the overall difficulty of the text or domain to ensure a wider range of values. Moreover, these participants may not be evaluating the type of effort being exerted during the task, relying instead on a different notion of what is effortful. Triangulation with retrospective verbalizations might help explain the patterns observed in the data.

As was the case in the correlations between perceived effort and technical effort, when perceived effort was compared to exerted effort by match type, again participants showed correlations in fuzzy and exact matches. Translators seem to perceive effort more readily in segments that they edited than those requiring translation. The data indicate that participants were more aware of the effort required to edit and modify proposed translations than they were of the effort required to furnish novel translation. The entire triadic bundle of problem recognition, solution proposal and solution resolution seems to be more apparent in these conditions. Translators may in part attribute effort to editing since problem recognition and solution evaluation are foregrounded and constantly in play, while in translation, these steps may only occur intermittently.

Inferential and correlational statistics like those presented and discussed in chapters 4 and 5 provide a better understanding of the type of effort exerted when translators undertake computer-assisted translation. The data also provide insight into the relationships that may exist between perception of effort and actual behavior. Chapter 6 will discuss the implications of these findings.

CHAPTER 6

CONCLUSION

The findings of the experimental study findings reported in this dissertation highlight the difference in the translator's cognitive effort when working with translation memory tools. It should be noted that these tools change the nature of the translation process. Translators that are working without a TM must translate a text from scratch and self-revise their work. When translators work with TM, however, the previously described behavior only occurs with no-match segments; in the remaining segments, t the data presented in this dissertation indicate that this change in the nature of the translation process in turn changes the cognitive effort exerted by translators.

More specifically, translators manifest a significant difference between the behavior they exhibit when editing and translating using a translation memory. Using the average pause ratio as a measure of cognitive effort, the data reveal that editing exact matches requires less cognitive effort than translating a segment from scratch. The data analysis, however, does not show a significant difference in effort when comparing fuzzy matches to either exact or new matches. Order effects ostensibly play a role in the lack of differentiation in the cognitive effort exerted by translators when translating segments that had either a fuzzy match or no match presented. In addition, editing behavior exhibited by all participants suggests that translators perceive no meaningful difference in the acceptability of fuzzy matches compared to exact matches, since participants systematically edited both match types. Exact matches were regularly edited by translators, and were not observed to require less cognitive effort as originally hypothesized. Further work is required to elucidate possible differences between the amount of effort participants exerted when working with fuzzy and exact matches.

Some participants showed strong correlations between the perceived effort required to translate or edit segments and the technical or temporal measures of effort. The majority of participants who conflated keystrokes, mouse clicks, or segment time with their perception of effort also showed a correlation with the actual effort exerted. These participants are likely assessing the cognitive effort required to edit segments and are simultaneously aware of the amount of technical intervention necessary to adapt the proposed translations. Translators who did not show these correlations may not have found the translation task particularly difficult. That is to say, in the translator’s experience, the task was not challenging and as a result, their data are clustered too tightly around higher scores. This clustering may impede the observation of any relationship that could exist between the participant’s perception of effort and the amount of effort exerted.

Implications

Pedagogy

Changes in cognitive effort and translator behavior in this study demonstrate the importance of including editing and computer-assisted translation in translator training curricula. Let us consider each in turn.

A marked decrease in effort between new and exact matches highlights a difference in the match types, and ultimately the two tasks being performed in each segment. In the former, translators provide an acceptable translation of the source text. Translation is a writing process with the source text serving as a point of departure for the new text. The triadic metacognitive bundle of problem recognition, solution proposal, and solution evaluation occurs as the translator works with the source language segment. Translators do not encounter interference from a proposed translation, and subsequently can translate the segment based on their understanding of the text, their experience, and their personal translation style.

In the latter, translators must evaluate the acceptability of the proposed translation. Cross-language evaluation first requires translators to understand the source text, and then verify that the translation adequately conveys the content of the original text. Carl et al. (2011) suggest a difference in the type of source text comprehension when translators translate from scratch compared to when they post-edit statistical machine translation (SMT) output using eye-tracking data. In Carl et al.'s study, participants refer to the source text with greater frequency and with longer fixations when translating the text than when post-editing. The authors posit:

> Manual translation seems to imply a deeper understanding of the ST, requiring more effort and thus longer fixations, whereas in post-editing, the ST is consulted frequently but briefly in order to check that the SMT output is an accurate and/or adequate reproduction of the ST. (Carl et al. 2011:140)

This dissertation's findings corroborate the difference in cognitive effort asserted by Carl et al. (2011). This dissertation's data show that translators exert effort differently in these two segment types, and therefore need to develop two complementary skill sets: translation and editing.

Differences in translation behavior observed in this dissertation also argue in favor of the development of these two skill sets. As outlined in chapters 4 and 5, participants exhibited distinct editing behaviors as they worked through the text. Some translators worked sequentially through the text, translating and editing a segment in its entirety before moving to another portion of the text. Others, in contrast, completed the translation and then systematically edited the text once the draft was complete. This behavior is in line with Carl's, Dragsted's, and Jakobsen's (2011) proposed classification of online, constant, and end revision. The authors define these three categories of revision as follows:

- *online revision*: the translator revises the text during the drafting phase;
- *end revision*: the translator spends 20 per cent or more of his/her time on end revision

- *constant revision*: the translator spends more than 20 per cent of translation time on end revision, but at the same time makes a large number (above average) of online revisions (Carl, Dragsted, and Jakobsen 2001: n.p.)

Regardless of the overarching editing behavior, all of the participants edited exact and fuzzy segments. Since translation memory tools are predicated on translators only having to edit fuzzy matches, this extra editing that occurs in exact matches arguably demonstrates over-editing by participants. Exact matches, by definition, are identical to previously translated segments and thus should in theory require no additional revision. If true, translators should not be introducing changes in these segments. Moreover, the fact that fuzzy and exact matches were treated similarly by the participants suggests that translators do not differentiate between the two segment types.

Translation trainers, then, should align their pedagogical practices with this type of empirical research to develop the requisite competences in translation students. Since the participants were observed to treat fuzzy and exact matches identically and over-edit segments, post-editing guidelines should be included in translation curricula to better prepare students to work with translation memory matches. Mossop (2001: 89) notes that revising both machine translation output and translation memory matches should be done "by making the minimum number of changes needed to create an acceptable translation." Training students to intervene as minimally as possible would be one way to address the over-editing exhibited by participants in this study.

The type of linguistic and structural edits that need to be introduced in segments also should be addressed in translation curricula. One example of how to include the types of changes commonly required when editing can be drawn from research on the inclusion of machine translation post-editing in university programs. Depraetere (2010: n.p.) uses a corpus of post-edited texts to identify specific machine translation errors that need to be highlighted for learners, “such as calque, mistranslation and untranslated ST.” This research allows trainers to target specific issues that often occur in machine translation. Further analysis of editing behavior is required to suggest specific editing strategies to be taught; however, it would be beneficial to replicate this research to determine the editing strategies required when working with translation memory matches. This research can then inform pedagogical practices in order to train students to best identify potential problem areas.

The marked difference between the editing and translation tasks and the observed tendency of participants to over-edit offers compelling evidence that editing warrants inclusion in translation curricula. Translation students must be aware of the context in which translation memory tools are used and the impact that their use has on the translation itself. Jiménez-Crespo (2013:53) highlights the work of several scholars (e.g., Bowker and Mossop) on TM technology’s impact on the text, such as a decrease in terminological coherence or the naturalness of the TL text. By drawing attention to this influence, translation students will be better equipped to correctly use these tools. Moreover, training should extend beyond project team members to encompass project managers in order to better integrate translation memory tools into the translation workflow.

Moreover, the task switching required in a segmented translation memory environment must also be highlighted. As has been previously discussed, the effort required in segments that require editing and in segments that must be translated differs significantly. Thus, translators must be trained to work with these different match types, and be able to consistently alternate between them. As observed in this study, translators occasionally returned to previously translated and edit segments. This behavior shows awareness of the text as a whole and potentially of the relationship between segments. Students must learn to address macro-level textual features, such as cohesive markers, when working in a segmented environment.

How editing skills are implemented in a translation curriculum depends on the structure that is currently in place. Many translation programs in the U.S. are certificate programs, which limits the number of credit hours that can be dedicated to a specific subject or the number of contact hours that can be devoted to a given topic. Master's level programs are typically two years in length, but are still subject to time constraints and credit restrictions when integrating new content into the program. Curricular changes to incorporate new classes can take significant lead time to implement, and adding courses to an already full M.A. program can prove difficult. In light of these curricular and time challenges, editing of translation memory matches or post-editing of machine translation can be integrated in several ways. Translation-specific editing courses specific to translators are a natural vehicle for the introduction of this type of work. Likewise, this type of material can be included in courses devoted to computer-assisted translation tools.

Perhaps the most effective way to develop editing skills in students would be to incorporate editing tasks into several courses that are closely articulated. Rather than treating CAT tool, localization and technology courses as independent areas of study that are not related to translation and editing courses, program administrators might instead consider how they are related. If specialized translation courses, such as technical or medical translation, were to incorporate CAT tools, students would learn to translate and edit these types of segments in a setting that reproduces real-world constraints. Ultimately the extent to which these tools can be integrated in these specialized translation courses depends on the program's infrastructure and the instructors' technological competence; however, the benefits of including CAT tools in this way are numerous.

Research Design and Methodology

The experimental study demonstrates the possibility of online process data collection. As noted in chapter 3, one of the challenges researchers face when conducting process-oriented research is a limited participant pool. Translators are often subcontractors and therefore have the flexibility of working remotely from anywhere in the world. In contrast, researchers cannot test in a laboratory setting large numbers of translators with a similar profile, but instead must rely on those translators who are available and local.[33] One way that researchers have obtained a larger number of participants is to study students enrolled in translation programs. Research on novice translators and the development of expertise has greatly benefited from their inclusion in research studies. Understanding the progression from novice to expert, though, requires investigation of the behavior of translators who have more experience than those enrolled in university programs.

Online data collection offers one possible way to address this issue. Research methods typically used in laboratory settings, such as keystroke logging, provide a wealth of information concerning the translation process and translator behavior. Adapting this methodology to an online environment requires several modifications to the research protocol. First, it is necessary to use software that can be deployed using a web server. Moreover, the researcher needs to consider issues of confidentiality and data security to ensure an appropriate level of both. Physical laboratory conditions can be easier to control than their online counterparts, and as a result, the researcher must try to control as many variables as possible to ensure the study results do not have any confounds.

This dissertation clearly demonstrates that these modifications can be made successfully to collect process data from geographically distributed translators. The researcher took care to minimize participant variables that could affect this particular study's results; however, future projects could include a greater number of translators

[33] As of 2012, the U.S. Bureau of Labor Statistics counted 63,600 people working in the U.S. as a translator or interpreter (Department of Labor 2014). If we take into account the diversity of languages, language pairs and direction, the difference between translation and interpreting, and areas of specialization, technology and years of experience, the number of translators that could qualify for a given study is relatively small.

with different language combinations or years of experience. Research questions that require keystroke logging can vary, but the protocol for collecting data online can remain constant. This methodology is well suited to a wide range of research designs that use keystroke logging. Ultimately, this experimental study shows that keystroke logging is a viable solution for online process data collection.

Tool Design

The main focus of this dissertation was the cognitive effort required to translate or edit three different types of segments — fuzzy match, exact match, and no match — and whether the translator's perception of effort coincided with exerted effort. The way in which translators perceive and exert effort is of particular importance to computer-assisted translation tool designers, since these measures indicate navigational preferences. For example, translators who conflate mouse clicks — a measure of technical effort — with exerted effort may prefer a translation memory system that does not require significant use of the mouse. Conversely, participants who associate keystrokes with effort may benefit from support at the subsegment level.

One way that support can be provided within the segment is interactive text prediction (ITP) to aid translators in their work. This type of interactive assistance is described in Alabau, Bonk, Buck et al. (2013) and enhances computer-assisted translation tool functionality by not only proposing a full segment for editing, but also by providing a new translation words or phrase proposals as edits are introduced.[34] Alabau et al. (2013) note that large variations exist between translators, and that translators often differ in the type of support that they prefer. That said, interactive text prediction was one type of support that participants found most beneficial.

This dissertation highlights distinct behavioral profiles revealed by the data correspond to translator typologies and how effort is perceived when editing proposed translation segments. Some translators associated technical effort with exerted effort, while others associated temporal effort. Further work is needed to understand the type of support that each of these translator profiles might require; however, the findings from this dissertation provide a foundation from which this type of work could be done. Knowing that technical and temporal effort are perceived differently by translators could affect software usability and human-computer interaction. Moreover, unique translator profiles emphasize the importance of having multiple ways to support a translator in his or her work.

Economics of Translation

As noted previously, the design and value proposition of TM tools are predicated on the assumption that exact matches require no revision; nevertheless, participant behavior clearly indicates substantial editing occurring in both fuzzy and exact matches.

---

[34] This research is part of a large-scale project, Cognitive Analysis and Statistical Methods for Advanced Computer Aided Translation (CASMACAT). All of the publications and research that has been conducted thus far are available at www.casmacat.eu/.

This is problematic, since compensation models in the language industry provide discounts for fuzzy and exact matches. For example, many translation providers charge for fuzzy matches using a sliding rate scale. The underlying assumption for this compensation model is that the rate of pay is proportional to the work effort required of the translator. These prices may be differentiated depending on the match level; a ninety percent match may be charged at a lesser rate than perhaps a seventy percent match. Exact matches are often the match type priced at the lowest rate, since these presumably do not require the translator to do any more than verify the translation.

If, however, translators are actively editing these matches and treating them as if they were fuzzy matches, this compensation model breaks down. Rather than exact matches being solely a verification task, the data suggest that participants perceive exact matches as similar to any other fuzzy match that has been presented. In this case, deeper discounts for exact matches are nonsensical. Moreover, if exact matches are inappropriate in the new translation context or contain errors, then translators may have to exert additional effort to adapt or correct these proposed translation solutions. This scenario is possibly more problematic for compensation models, since the level of intervention may supersede that of over-editing.

The researcher acknowledges that some CAT tools are able to present exact matches in a locked or inalterable state, such that translators can see the source-language and target segments of the exact-match translation units in their entirety but cannot make any changes to the translations. It is important that locked segments display even if only in read-only mode so that translators can view materials requiring translation or revision in the larger context of the document as a whole. Otherwise, translators will not have sufficient contextual information to translate or edit no-match and fuzzy match segments. In this case, perhaps exact matches could continue being priced using a sliding rate scale, although the rationale for this type of compensation model is suspect.

Furthermore, the similarity in treatment of fuzzy and exact matches calls into question the value being added by the translator. Over-editing of acceptable segments may not add value to the translation, and instead represents a duplication of effort. Translation memory serves as a way to reuse previously translated material, and editing these segments may be counterproductive. In addition, translation buyers may not want to compensate translators for duplicating effort that does not necessarily need to be exerted. Therefore, edits need to actively add value to the product and translators should be able to articulate this added value. That said, one should keep in mind that exact matches are determined by a comparison with the source text segment, and the match level does not take into account the quality of the proposed translation. The observed participant behavior of over-editing may in fact be an indication of potential deficiencies in presumably acceptable translations. Further research is necessary with vetted exact matches to determine whether the observed behavior is the result of defective exact match proposals.

The possibility of exact matches not being optimal translations raises another important consideration; the importance of maintaining the quality of translation memories. As stated previously, the observed behavior of participants over-editing exact matches may be indicative of sub-optimal translations being stored and then presented to the translator by the TM tool. If these translations have been stored and ultimately require additional intervention on the part of the translator during the translation task, perhaps translation agencies and translators need to place greater emphasis on the quality of translations stored in the TM. In doing so, companies could limit over-editing and optimize the translation workflow by ensuring translators can focus on segments that require editing.

Moreover, additional study is needed to determine if the translator's level of expertise affects the overall tolerance of different translations for a single source segment. In this dissertation, all of the participants have a comparable amount of experience (4–7 years) and exhibited similar editing behavior of exact and fuzzy matches. Further study is necessary to investigate if novice or expert translators exhibit the same over-editing tendencies, or if perhaps experts are less likely to intervene in the proposed target language segments than their less experienced counterparts.

Cognitive Model

The previous discussion of over-editing can be tied back to the cognitive model of uncertainty management proposed by Angelone (2010). The triadic bundle of problem recognition, solution proposal, and solution evaluation help elucidate the type of behavior shown by participants. While further research is necessary to corroborate these hypotheses, the data presented in this dissertation suggest that problem recognition is responsible in part for over-editing. The participants in this study seem to identify problems in both fuzzy and exact matches equally. The amount of effort exerted, as measured by the average pause ratio (APR), is also roughly equal in these two match types. The overall amount of exerted effort in no-match segments though is considerably greater than that exerted in fuzzy and exact match segments. Consequently, while problem recognition seems to occur with similar frequency in fuzzy and exact match segments, this cognitive behavior ostensibly requires less effort than no-match segments. If so, the difference in cognitive effort exerted must lie in the solution proposal and solution evaluation stages.

As noted previously, the exertion of cognitive effort differs between no-match segments and the comparable fuzzy and exact matches. In no-match segments, problem recognition is specific to the translation task and self-revision. In contrast, fuzzy and exact match problem recognition is associated with identifying deviations between the source and target segments. If the solution proposal and solution evaluation stages of the translator's cognitive effort are the stages in which the exerted cognitive effort differs, then perhaps the segment type is causing this change. Data from this dissertation suggest that the proposal of solutions and evaluating these proposals is less effortful when translators do not have to evaluate their own work. Instead, translators exert less effort when editing a translation that has been proposed by a translation memory, regardless of the match level. This finding is curious, since correlations were found between the translator's perception of effort of editing fuzzy and exact matches and the amount of effort that was actually exerted.

One potential explanation for cognitive effort being greater to propose and evaluate solutions in novel translation than in editing a match type is related to the nature of the solution proposal stage. The generation of a new solution is a writing process, and ultimately the creation of new text may be more effortful than adapting an existing translation. That said, the perception of effort being greater to edit an existing translation may indicate that participants experience interference from the target language proposal. This interference is not necessarily more effortful and does not present as such in the data, but translators perceive the conflict between the translation proposal and their own rendition as being more difficult to overcome than creating a new translation.

As noted previously, these hypotheses require additional investigation to better understand the effort being exerted by translators that are working with translation memory matches. Nevertheless, the data provide evidence than the solution proposal and solution evaluation stages require greater scrutiny in an effort to understand how the use of translation memory affects the cognitive behavior of the translator.

Limitations of the Study

As with all studies, there are limitations as to how much the results presented here are generalizable. Indeed, the findings strongly suggest differences in translator's cognitive effort when working with new and exact matches. Likewise, the strong correlations established in participants are indicative of a conflation of different types of effort (technical, temporal, cognitive) and perceived effort. Nevertheless, the participant pool still remains relatively small. The fact that participants were quite similar mitigates the small size of the participant group in part, and allows very specific indicators of effort to be measured. Nevertheless, it is difficult to argue that the results presented for these nine participants will hold for all translators, in all language pairs and directions, at all levels of experience.

The goal of the study, however, was not to find results that will hold across all of these divisions. Instead, the experiment aimed to investigate effort in this specific group of participants. This demographic of participants, in particular translators with 4 to 7 years of experience, has been largely understudied and merits greater attention. The dissertation begins to fill this gap in the literature.

Order effects are also another limitation in the study. As discussed previously, order effects in the experimental task potentially muddle any differences that could otherwise be observed between fuzzy matches and the remaining two match types. These effects, however, are inherent in this type of stimulus. Each segment is presented sequentially, and therefore, there is invariably some potential for participants to be affected by what precedes and what follows a given segment. Had the experimental text been such that segments could be presented non-sequentially, in a truly random order, segments could be presented in a more counterbalanced manner, which in turn would minimize any potential order effects.

Another limitation is the unfamiliarity of the tool to participants. Each translator that was part of the study had 4 to 7 years of experience, and also had experience using CAT tools. The tool used to record translator behavior, TransCenter, is not a commercially available tool, nor is it one that professional translators would use to translate or edit. Consequently, translator behavior may have been influenced by a lack of familiarity, and subsequently changed as translators became more familiar with the tool during the time required to complete the task. This limitation was addressed somewhat by the introduction to the software provided by the researcher to the software. Instructions were provided in the email that included each participant's log-in credentials, and additional screenshots and information related to the study were available as web help documentation. Participants were instructed to read this information online prior to commencing the task, and they were advised that they could pause the task to reference the materials at any time. Nevertheless, future iterations of the study may require participants to complete a sample before beginning the experimental task, or provide some kind of online training.

A final limitation in this study is that participants were not compensated. Translators were encouraged to complete the task as they would any professional translation; however, their motivation may have differed given the lack of remuneration. Participants self-selected to be part of the study and presumably see the benefit in taking part in research. That said, several participants were very busy and did not complete the experiment study in a timely manner after initially agreeing to participate since their professional work naturally took precedence.

Future Directions

Several scholars in the field have taken up research on translation memory and its impact on the translator's cognition. This dissertation follows in their wake, and has already noted several areas in which additional work is necessary. The limitations outlined above suggest several possible avenues of future research. For example, it is necessarily to more closely examine order effects of match types for us to be able to more clearly understand the impact that proposed translations have on the text (similar to those suggested by Mossop 2006 and Jiménez-Crespo 2013) and translator behavior. Likewise, expanded participant pools would greatly benefit the generalizability of the present study.

Additional methodological work is necessary to develop and test online data collection methods to help triangulate findings with keystroke measures. This type of work will help expand the repertoire of data elicitation methodologies available to translation scholars in their effort to understand the translation process. Such work may triangulate the present study using retrospective verbalizations or directed interviews. Likewise, mouse-tracking may be a potential avenue for additional data. Eye-tracking studies are a logical progression of this project, although this methodology may prove difficult to implement in an online setting.

Linguistic analysis of the edits made by translators in this study will be beneficial to our understanding of how cognitive effort is being applied. Findings of this type of study would help develop pedagogical interventions that could better train students to edit proposed translations in CAT tools. Lacruz, Shreve, and Angelone's (2012) metric of complete editing events will also elucidate which types of segments may be more difficult for translators to edit as they progress through the text. Order effects may also have an impact on any linguistic analysis, and further work is required to elucidate the impact that order effects may have on the translation.

Finally, replication of the present study using different participant variables would prove useful. Similar results would help generalize the findings presented here, and may illuminate differences in cognitive effort and translator behavior across languages and experience levels. Expertise studies would greatly benefit from replication, since little research has been conducted to date that focuses on the behavior of translators with 4 to 7 years of experience is minimal. The development of expertise occurs over time, and the documentation of translator behavior between novice and expert provides data that could provide the starting point for a longitudinal study. This type of replication may ultimately allow the researcher to refine models of cognitive effort and better generalize translation behavior in a computer-assisted translation environment.

REFERENCES

Ahmad, Khurshid and Margaret Rogers. "Corpus-Related Applications: Corpus Linguistics and Terminology Extraction." In *Handbook of Terminology Management: Application-Oriented Terminology Management. Vol. 2*, edited by Sue Ellen Wright and Gerhard Budin, 725–760. Philadelphia: John Benjamins, 2001.

Alabau, Vicent, Ragnar Bonk, Christian Buck, Michael Carl, Francisco Casacuberta, Mercedes García-Martínez, Jesús González-Rubio, Philipp Koehn, Luis A. Leiva, Bartolomé Mesa-Lao, Hervé Saint-Amand, Chara Tsoukala, Germán Sanchis-Trilles and Daniel Ortiz-Martínez. "Advanced Computer Aided Translation with a Web-Based Workbench." In *Proceedings of MT Summit XIV on Post-editing Technology and Practice*, edited by Sharon O'Brien, Michel Simard and Lucia Specia. European Association for Machine Translation: Allschwil, Switzerland, 55–62. 2013.

Alabau, Vincent, Jesús González-Rubio, Luis A. Leiva, Daniel Ortiz-Martínez, Germán Sanchis-Trilles, Francisco Casacuberta, Bartolomé Mesa-Lao, Ragnar Bonk, Michael Carl and Mercedes García-Martínez. "User Evaluation of Advanced Interaction Features for a Computer-Assisted Translation Workbench." In

Machine Translation Summit XIV: Main Conference Proceedings Proceedings of the XIV Machine Translation Summit, edited by Andy Way, Khalil Sima'an, Mikel L. Forcada, Daniel Grasmick, and Heidi Depraetere. European Association for Machine Translation: Allschwil, Switzerland, 361–368. 2013.

Alcina, Amparo. "Translation Technologies: Scope, Tools and Resources." *Target* 20, no. 1 (2008): 79–102.

Almeida, Giselle de, and Sharon O'Brien. "Analysing Post-Editing Performance: Correlation with Years of Translation Experience." In *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, edited by François Yvon and Viggo Hansen. May 27–28, 2010. n.p. Accessed March 2, 2014. http://bit.ly/1g2DVG0.

Alves, Fabio, Adriana Pagano, Stella Neumann, Erich Steiner, and Silvia Hansen-Schirra. "Translation Units and Grammatical Shifts: Toward an Integration of Product- and Process-based Translation Research." In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 109–142. Philadelphia: John Benjamins, 2010.

Alves, Fabio, and Daniel Couto Vale. "Probing the Unit of Translation in Time: Aspects of the Design and Development of a Web Application for Storing, Annotating, and Querying Translation Process Data." *Across Languages and Cultures* 10, no. 2 (2009): 251–273.

Angelone, Erik. *The Conceptualization and Integration of an e-Collocation Trainer Methods of Empirical, Translation Based Collocation Research*. Trier: WVT Wissenschaftlicher Verlag, 2007.

Angelone, Erik. “Uncertainty, Uncertainty Management, and Metacognitive Problem Solving in the Translation Task.” In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 17–39. Philadelphia: John Benjamins, 2010.

Arnold, Douglas, Lorna Balkan, Siety Meijer, R. Lee Humphreys, Louisa Sadler. *Machine Translation: An Introductory Guide*. London: Blackwell, 1994.

ASTM International. “ASTM F 2575 - 06: Standard Guide for Quality Assurance in Translation.” 2006.

Austermühl, Frank. *Electronic Tools for Translators.* Manchester: St. Jerome Publishing, 2001.

Baddeley, Alan, and Graham Hitch. “Verbal Reasoning and Working Memory.” *Quarterly Journal of Experimental Psychology* 28 (1976): 603–621.

Baddeley, Alan, and Graham Hitch. “Working Memory.” *Psychology of Learning and Motivation* 8 (1974): 47–89.

Balling, Laura. “A Brief Introduction to Regression Designs and Mixed-Effects Modelling By A Recent Convert.” In *Looking at Eyes: Eye Tracking Studies of Reading and Translation Processing*, edited by Susanne Göpferich, Arnt Lykke Jakobsen and Inger Mees. Copenhagen: Samfunslitteratur, 175–192.

Beaton, Adriana, and Gabriela Contreras. "Sharing the Continental Airlines and SDL Post-editing Experience." *AMTA 2010: the Ninth Conference of the Association for Machine Translation in the Americas*, Denver, Colorado, October 31 – November 4, 2010. Accessed March 1, 2014. http://bit.ly/1pKhAVy.

Bowker, Lynne. *Computer-Aided Translation Technology*. Ottawa: University of Ottawa Press, 2002.

Bowker, Lynne and Jennifer Pearson. *Working with Specialized Language: A Practical Guide to Using Corpora*. New York: Routledge, 2002.

Buendía-Castro, Miriam and Clara Inés López-Rodríguez. "The Web for Corpus and the Web as Corpus in Translator Training." *New Voices in Translation Studies* 10, (2013): 54–71.

Bureau of Labor Statistics, U.S. Department of Labor. *Occupational Outlook Handbook, 2014–15 Edition*, Interpreters and Translators. Accessed March 1, 2014. http://1.usa.gov/NhlAhy.

Butterworth, Brian. "Evidence from Pauses in Speech." *Language Production: Speech and Talk. Vol.1*, edited by Brian Butterworth, 155–176. London: Academic Press, 1980.

Campbell, Stuart. "Choice Network Analysis in Translation Research." In *Intercultural Faultlines: Research Models in Translation Studies: Textual and Cognitve Aspects*, edited by Maeve Olohan, 29–42. Manchester: St. Jerome Publishing, 2000.

Carl, Michael, Barbara Dragsted, and Arnt Lykke Jakobsen. “On the Systematicity of Human Translation Processes.” In *Proceedings of Translation Careers and Technologies: Convergence Points for the Future (Tralogy 2011)*, Paris, March 3 2011. n.p. Accessed March 7, 2014. http://bit.ly/1hWWkro.

Carl, Michael, Barbara Dragsted, Jakob Elming, Daniel Hardt, and Arnt Lyyke Jakobsen. “The Process of Post-Editing: A Case Study.” In *Proceedings of the 8th International NLPSC Workshop*, edited by Bernadette Sharp, Michael Zock, Michael Carl, Arnt Lykke Jakobsen, 131–142. Frederiksberg: Samfundslitteratur, 2011. Accessed March 1, 2014. http://bit.ly/1pL51JI.

Cowan, Nelson. *Attention and Memory: An Integrated Framework*. no. 26. Oxford University Press on Demand, 1995.

DePalma, Donald A., and Robert G. Stewart. *Trends in Translation Pricing: Falling Rates Accompany Changes in Economy and Buying Behaviors.* Lowell, MA: Common Sense Advisory, 2012.

DePalma, Donald A., Robert G. Stewart, and Belinda Whittaker. *Translation and Localization Pricing: A Comprehensive Study of What Language Services Cost.* Lowell, MA: Common Sense Advisory, 2010.

Denkowski, Michael, and Alan Lavie. “TransCenter: Web-Based Translation Research Suite.” A*MTA 2012 Workshop on Post-Editing Technology and Practice Demo Session*. 2012.

Depraetere, Ilse. “What Counts as Useful Advice in a University Post-Editing Training Context? Report on a Case Study.” In *EAMT 2010: Proceedings of the 14th*

Annual Conference of the European Association for Machine Translation, edited by Viggo Hansen and François Yvon. Saint-Raphaël, France, May 27–28, 2010. n.p. Accessed March 1, 2014. http://bit.ly/1ofMTnM.

Dragsted, Barbara. “Computer-aided Translation as a Distributed Cognitive Task.” In *Cognition Distributed: How Cognitive Technology Extends our Minds*, edited by Itiel E. Dror and Stevan Harnad, 237–256. Philadelphia: John Benjamins, 2008.

Dragsted, Barbara. *Segmentation in Translation and Translation Memory Systems: An Empirical Investigation of Cognitive Segmentation and Effects of Integrating a TM-System into the Translation Process.* Copenhagen: Samfundslitteratur, 2004.

Dragsted, Barbara. “Segmentation in Translation: Differences Across Levels of Expertise and Difficulty.” *Target* 17, no. 1 (2005): 49–70.

Dragsted, Barbara, and Inge Gorm Hansen. “Comprehension and Production in Translation: A Pilot Study on Segmentation and the Coordination of Reading and Writing Processes.” In *Looking at Eyes*, edited by Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees, 9–30. Copenhagen: Samfundsliteratur, 2008.

Dunne, Keiran J. “Computer-Assisted Translation.” In *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle, 839–843. Blackwell Publishing, 2013.

EMC. “The 2011 IDC Digital Universe Study.” http://www.emc.com/collateral/about/news/idc-emc-digital-universe-2011-infographic.pdf. 2011. Accessed 4 February 2014.

Ericsson, K. Anders, and Walter Kintsch. “Long-term Working Memory.” *Psychological review* 102, no. 2 (1995): 211.

Esselink, Bert. *A Practical Guide to Localization*. Philadelphia: John Benjamins, 2000.

Guerberof Arenas, Ana. "Productivity and Quality in the Post-Editing of Outputs from Translation Memories and Machine Translation." *The International Journal of Localisation* 7, no. 1 (2009): 11–21.

Hartley, Tony. "Technology and Translation." In *The Routledge Companion to Translation Studies*, edited by Jeremy Munday, 106–127. New York: Routledge, 2009.

Heiman, Gary W. *Understanding Research Methods and Statistics: An Integrated Introduction for Psychology*. *Second Edition*. New York: Houghton Mifflin, 2001.

Hutchins, John. "Machine Translation: A Concise History." *Journal of Translation Studies* 13, no. 1 (2010): 29–70.

Hutchins, John. "The Origin of the Translator's Workstation." *Machine Translation* 13, no. 4 (1998): 287–307.

IDC. "The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011." http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf. March 2008. Accessed 4 February 2014.

Immonen, Sini. "Translation as a Writing Process: Pauses in Translation versus Monolingual Text Production." *Target* 18, no. 2 (2006): 313–335.

International Organization for Standardization. *ISO 704: 2009. Terminology Work – Principles and Methods*. Geneva: ISO, 2009.

International Organization for Standardization. *ISO 12620: 1999. Computer Applications in Terminology – Data Categories*. Geneva: ISO, 1999.

Jääskeläinen, Riitta. “Think-aloud Protocol Studies into Translation: An Annotated Bibliography.” *Target* 14, no. 1 (2002): 107–136.

Jakobsen, Arnt Lykke. “Logging Time Delay in Translation, LSP Texts and the Translation Process.” *Copenhagen Working Papers*. (1998): 73–101.

Jakobsen, Arnt Lykke. “Logging Target Text Production with Translog.” In *Probing the Process in Translation: Methods and Results*, edited by Gyde Hansen, 9–20. Copenhagen: Samfundslitteratur, 1999.

Jakobsen, Arnt Lykke. “Tracking Translators’ Keystrokes and Eye Movements with Translog.” In *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, edited by Cecilia Alvstad, Adelina Hild and Elisabet Tiselius, 37–55. Philadelphia: John Benjamins, 2011.

Jiménez-Crespo, Miguel. *Translation and Web Localization*. New York: Routledge, 2013.

Johnson, Keith. *Quantitative Methods in Linguistics*. Oxford: Blackwell Publishing, 2008.

Kelly, Nataly, Donald A. DePalma, and Robert G. Stewart. *The Language Services Market: 2012*. Lowell, MA: Common Sense Advisory, May 2012.

Kenny, Dorothy. “Electronic Tools and Resources for Translators.” In *The Oxford Handbook of Translation Studies*, edited by Kirsten Malmkjær and Kevin Windle, 455–472. Oxford: Oxford University Press, 2011.

Kilgray Translation Technologies. *Translators*. http://kilgray.com/solutions/user-group/translators. Accessed 22 January 2013.

Krings, Hans P. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes.* Edited by Geoffrey S. Koby. Kent, OH: Kent State University Press, 2001.

Koponen, Maarit. “Comparing Human Perceptions of Post-editing Effort with Post-editing Operations.” *Proceedings of the 7th Workshop on Statistical Machine Translation*, 181–190, Montreal, Canada, June 7–8, 2012. Accessed March 1, 2014. http://bit.ly/1pK9hsP.

Krings, Hans P. *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes.* Edited by Geoffrey S. Koby. Kent, OH: Kent State University Press, 2001.

Lacruz, Isabel, Gregory M. Shreve, and Erik Angelone. “Average Pause Ratio as an Indicator of Cognitive Effort in Post-Editing: A Case Study.” *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice.* San Diego, CA, 2012.

Lacruz, Isabel and Gregory M. Shreve. “Pauses and Cognitive Effort in Post-Editing.” In *Post-Editing of Machine Translation: Processes and Applications*, edited by Sharon O'Brien, Laura Winther Balling, Michael Carl, Michel Simard and Lucia Specia. 244–272. Cambridge: Cambridge Scholars Publishing, 2014.

Mauthe, Andreas, and Peter Thomas. *Professional Content Management Systems: Handling Digital Media Assets*. West Sussex, England: Wiley, 2004.

Mossop, Brian. *Revising and Editing for Translators*. Manchester: St. Jerome Publishing, 2001.

O'Brien, Sharon. "An Empirical Investigation of Temporal and Technical Post-Editing Effort." *Translation and Intpreting Studies*, 2007: 83–136.

O'Brien, Sharon. "Eye-Tracking and Translation Memory Matches." *Perspectives: Studies in Translatology* 14, no. 3 (2006a): 185–205.

O'Brien, Sharon. "Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output." *Across Languages and Cultures* 7, no. 1 (2006b): 1–21.

O'Brien, Sharon. "Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability." *Machine Translation* 19, no. 1 (2005): 37–58.

O'Brien, Sharon. "Processing Fuzzy Matches in Translation Memory Tools: an Eye Tracking Analysis." In *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*, edited by Susanne Göpferich, Arnt Lykke Jakobsen, and Inger M. Mees, 79–102. Copenhagen: Samfundslitteratur, 2008.

Olohan, Maeve. *Introducing Corpora in Translation Studies*. New York: Routledge, 2004.

Ping, Ke. "Machine Translation." In *Routledge Encyclopedia of Translation Studies. 2nd Edition*, edited by Mona Baker and Gabriela Saldanha. 162–169. Routledge: New York, 2009.

Proctor, Robert W., and Kim-Phuong L. Vu. "Human Information Processing: An Overview for Human-Computer Interaction." In *The Human-Computer*

Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, edited by Julie A. Jacko, 21–40. Boca Raton, FL: CRC Press, 2012.

Quah, C. K. *Translation and Technology*. New York: Palgrave Macmillan, 2006.

Saldanha, Gabriela, and Sharon O'Brien. *Research Methodologies in Translation Studies*. Manchester, UK: St. Jerome, 2013.

Samson, Richard. "Computer-Assisted Translation." In *Training for the New Millenium*, edited by Martha Tennent, 101–126. Philadelphia: John Benjamins, 2005.

Schmitz, Klaus-Dirk. "Criteria for Evaluating Terminology Database Management Programs." In *Handbook of Terminology Management: Application-Oriented Terminology Management. Vol. 2*, edited by Sue Ellen Wright and Gerhard Budin, 539–551. Philadelphia: John Benjamins, 2001.

SDL. *About PerfectMatch*. November 13, 2013. Accessed January 20, 2014. http://bit.ly/1dQwwbc.

SDL. *SDL Trados Studio 2011: Market-leading translation software*. http://www.sdl.com/products/sdl-trados-studio/. Accessed 22 January 2013.

Shreve, Gregory M., and Bruce J. Diamond. "Cognitive Processes in Translation and Interpreting: Critical Issues." In *Cognitive Processes in Translation and Interpreting*, edited by Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain, and Michael McBeath , 233–252. Thousand Oaks, CA: Sage Press, 1997.

Shreve, Gregory M., Isabel Lacruz, and Erik Angelone. "Sight Translation and Speech Disfluency: Performance Analysis as a Window to Cognitive Translation

Processes." In *Methods and Strategies of Process Research*, edited by Cecilia Alvstad, Adelina Hild, and Elisabet Tiselius, 93–120. Philadelphia: John Benjamins, 2011.

Spalink, Karin, Rachel Levy, and Carla Merrill. *The Level Edit ™ Post-Editing Process: A Tutorial for Post-Editors of Machine Translation Output*. Internationalization and Translation Services, 1997.

Sweller, John. "Human Cognitive Architecture." In *Handbook of Research on Educational Communications and Technology*, edited by Michael J. Spector, M. David Merrill, Jeroen van Merrienboer, and Marcy P. Driscoll, 369–381. New York: Routledge, 2007.

Sweller, John. "Cognitive Load during Problem Solving: Effects on Learning." *Cognitive Science* 12, no. 2 (1988): 257–285.

TAUS. "Machine Translation Post-Editing Guidelines." November 2010. https://evaluation.taus.net/resources/guidelines/post-editing/machine-translation-post-editing-guidelines. Accessed January 23, 2014.

Underwood, Nancy L., and Bart Jongejan. "Translatability Checker: A Tool to Help Decide Whether to Use MT." *Proceedings of MT Summit VIII: Machine Translation in the Information* Age, edited by Bente Maegaard. Santiago de Compostela, Spain, 2001. 363–368. Accessed March 1, 2014. http://bit.ly/1hu4w2W.

Waes, Luuk Van, and Peter Jan Schellens. “Writing Profiles: The Effect of the Writing Mode on Pausing and Revision Patterns of Experienced Writers.” *Journal of Pragmatics* 35 (2003): 829–853.

Wright, Sue Ellen, and Gerard Budin. *Handbook of Terminology Management: Basic Aspects of Terminology Management. Vol. 1*. Philadelphia: John Benjamins, 1997.

Wright, Sue Ellen, and Gerhard Budin. *Handbook of Terminology Management: Application-Oriented Terminology Management. Vol. 2*. Philadelphia: John Benjamins, 2001.

Zanettin, Federico. “Parallel corpora in translation studies: Issues in corpus design.” *Proceedings of LREC 2002 Workshop: Language Resources and Translation Work and Research*. Las Palmas, May 29–31, 2002. 10–14.

CELEBRATE CENTENNIAL
KENT STATE
UNIVERSITY
1910·2010

Computer-Assisted Translation: An Empirical Investigation of Cognitive Effort

Welcome to "Effort in Translation," a web-based experiment designed to better understand how people translate when using computer-assisted translation (CAT) tools. Before taking part in this study, please read the consent form below and click on the "I Agree" button at the bottom of the page if you understand the statements and freely consent to participate in the study.

Consent Form

This study is being conducted by Ph.D. candidate Christopher Mellinger and Dr. Keiran Dunne of Kent State University, and it has been approved by the Kent State University Institutional Review Board. No deception is involved, and the study involves no more than minimal risk to participants (i.e., the level of risk encountered in daily life).

Participation in the study typically takes no more than 90 minutes and is strictly anonymous. Participants will be asked to translate a text of approximately 400 words from Spanish to English without using any resources other than those provided by the computer-assisted translation system. The text will appear on the left-hand side of the screen and you will have the option of either translating the segment or editing the suggested translation that appears in the box on the right-hand side of the screen.

All responses will be treated with absolute confidentiality, and in no case will responses from individual participants be identified. Rather, all data will be pooled and published in aggregate form only. Any reference to a specific participant will be made using an anonymous name. Participants should be aware, however, that the experiment is not being run from a "secure" https server of the kind typically used to handle credit card transactions, so there is a small possibility that responses could be viewed by unauthorized third parties (e.g., computer hackers).

Participation in this study will not entail any risk to you beyond those risks encountered in everyday life, and no adverse reactions have been reported. Participation is voluntary and refusal to take part in the study involves no penalty or loss of benefits to which you are otherwise entitled. You may withdraw from the study at any time.

If you have further questions about this study or your rights, or if you wish to lodge a complaint or concern, you may contact the principal investigator, Dr. Keiran Dunne, or the co-investigator, Ph.D. candidate Christopher Mellinger, at (908) 399-9252; or the Kent State University Institutional Review Board, at (330) 672-2704.

If you are 18 years of age or older, understand the statements above, and freely consent to participate in the study, click on the "I Agree" button to begin the experiment.

I Agree | I Do Not Agree

Appendix B:

Survey – Computer-Assisted Translation: An Empirical Investigation of Cognitive Effort

Which of the following best describes your current professional status?

- Freelance translator
- In-house translator
- Other: ____________________

What are your working language combinations (select all that apply)?

- Spanish to English
- English to Spanish
- French to English
- English to French
- German to English
- English to German
- Portuguese to English
- English to Portuguese
- Other:___________________

What is your native language (your primary target language)?

- English
- Spanish
- French
- German
- Portuguese
- Other:_____________________

Where are you currently based (if you split your time throughout the year in more than one location, where are you located most often)?

____

What types of language services do you offer (select all that apply)?

- Translation
- Editing
- Proofreading
- Post-editing (of machine translation)

- Interpreting
- Other: __________________

If you offer both translation and interpreting services, in which area do you do more work?

- More translation than interpreting
- About as much translation as interpreting
- More interpreting than translation
- n/a

What computer-assisted translation tools do you use (select all that apply)?

- SDL Trados Studio
- SDL Trados Workbench
- MemoQ
- OmegaT
- Wordfast
- Wordbee
- Across
- Lingotek
- StarTransit
- Other: __________________

In approximately what percentage of your translation work do you use CAT tools?

- 0%
- 1-25%
- 26-50%
- 51-75%
- 76-100%

Approximately what percentage of your income is derived from translation work?

- Less than 30%
- 30-70%
- 70-100%

How many years have you been working as a Spanish-to-English translator?

- 0
- 1-3
- 4-7
- 8-10
- 11+

Do you have any formal training in translation?

- Yes
- No

If so, please describe your training:_______________

Do you have any formal training in editing?

- Yes
- No

If so, please describe your training:_______________

Do you have any formal training in post-editing?

- Yes
- No

If so, please describe your training:_______________

Appendix C:

Survey Recruitment Script

Hello,

My name is Chris Mellinger and I am a doctoral candidate at Kent State University. I am conducting a study that will help me collect data for my dissertation, which will examine effort in translation when using computer-assisted translation software. This study is being conducted by Dr. Keiran Dunne, principal investigator, and co-investigator, Ph.D. candidate Christopher Mellinger, and it has been approved by the Kent State University Institutional Review Board.

I am currently looking for Spanish-to-English translation professionals who receive some or all of their income from the language industry to participate in this study. The study consists of two parts. The first part is a survey that will be used to determine your eligibility to take part in the second half of the study. The survey questions are about your work as a professional translator, and should take approximately 5 minutes to complete.

Should you qualify, the second part of the study consists of translating a text of approximately 400 words from Spanish to English using a web-based computer-assisted translation tool and without using any external resources. This translation should take no longer than 90 minutes to complete.

If you are a Spanish-to-English translator who receives all or some of your income from the language industry, I would greatly appreciate 5 minutes of your time to complete this survey, and if you qualify, 90 minutes to complete the experiment. Please click on the following link (qualtrics link) (or cut and paste it onto your internet browser) to complete the survey.

Thank you very much in advance.

Best regards,

Chris Mellinger

Appendix D:

Experimental instrument – Source Text

En su breve paso por Buenos Aires, el primer ministro japonés Shinzo Abe se encargó de remarcar que su país, y particularmente Tokio, es “seguro” para los Juegos Olímpicos 2020 que acaban de ganar. Exultante por haber obtenido lo que vino a buscar, el mandatario aseguró que la crisis desatada en la planta de energía nuclear de Fukushima está absolutamente controlada. También, defendió los millonarios programas oficiales para detener las filtraciones y descontaminar el agua radiactiva.

Sostuvo que “el gobierno ha establecido una serie de medidas para reducir la dependencia de la energía nuclear ” en los próximos tres años, al tiempo que aplicarán “los más severos programas científicos de seguridad ” sobre las centrales atómicas.

En Japón, la tercera economía mundial, la provisión de energía nuclear es sólo una parte mínima de la que utiliza su potente desarrollo. Es más, de sus 54 centrales nucleares sólo dos están en funcionamiento. Las otras se encuentran paradas. El país se está abasteciendo de otro tipo de energía y promoviendo la conservación y la adopción de energías renovables.

El accidente en Fukushima le dio un golpe mortal a la energía nuclar. En marzo de 2011 la central atómica se vio afectada por el terremoto de 9 grados que sacudió y devastó al país, para inmediatamente sufrir el tsunami que arrasó con el sistema de refrigeración de los reactores. Esto provocó un recalentamiento que trajo innumerables problemas. Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar. El mes pasado el gobierno decidió encarar el problema por su cuenta, desplazando a la operadora privada de la central, TEPCO's, que no encontraba una solución. Para ello estableció una serie de medidas tendientes a frenar las filtraciones.

Desplegó un programa que incluye un sofisticado sistema de congelamiento del terreno alrededor de la planta, con tubos que llevan un potente refrigerante. También piensa establecer un proceso novedoso de descontaminación del agua radiactiva, que se viene almacenando peligrosamente en contenedores. Estos trabajos comenzarán inmediatamente y se espera que concluyan en dos años.

"Hemos realizado monitoreos en la región costera de Fukushima. Las normas de seguridad de Japón tanto para alimentos como para aguas tienen el nivel más estricto del mundo".

Abe sostuvo que "no ha habido problemas relacionados con la salud hasta ahora. No los hay en el presente y no los habrá en el futuro, lo afirmo de la forma más enfática".

Appendix E:

Experimental instrument – Target Text Versions with Markup

Version A

1	EXACT	En su breve paso por Buenos Aires, el primer ministro japonés Shinzo Abe se encargó de remarcar que su país, y particularmente Tokio, es “seguro” para los Juegos Olímpicos 2020 que acaban de ganar.	During a brief stop in Buenos Aires, Shinzo Abe, prime minister of Japan, stressed that his country, and in particular Tokyo, is "safe" for the 2020 Olympic Games that were recently awarded to the city.
2		Exultante por haber obtenido lo que vino a buscar, el mandatario aseguró que la crisis desatada en la planta de energía nuclear de Fukushima está absolutamente controlada.	
3	FUZZY	También, defendió los millonarios programas oficiales para detener las filtraciones y descontaminar el agua radiactiva.	He also defended the millions ~~of~~ in official programs to stop leaks and decontaminate radioactive ~~water~~ waste.
4		Sostuvo que “el gobierno ha establecido una serie de medidas para reducir la dependencia de la energía nuclear ” en los próximos tres años, al tiempo que aplicarán “los más severos programas científicos de seguridad ” sobre las centrales atómicas.	
5	EXACT	En Japón, la tercera economía mundial, la provisión de energía nuclear es sólo una parte mínima de la que utiliza su potente desarrollo.	In Japan, the world's third-largest economy, nuclear energy production is only a small part used by its strong development.
6	FUZZY	Es más, de sus 54 centrales nucleares sólo dos están en funcionamiento.	~~In fact,~~ O~~o~~nly two of the ~~54~~ 45 nuclear power plants are currently operational.
7		Las otras se encuentran paradas.	

8	FUZZY	El país se está abasteciendo de otro tipo de energía y promoviendo la conservación y la adopción de energías renovables.	~~The country~~ Japan is ~~supplied by~~ supplying another type of energy and ~~is~~ promotes~~ing~~ energy conservation and the adoption of renewable energy.
9	EXACT	El accidente en Fukushima le dio un golpe mortal a la energía nuclar.	The Fukushima accident dealt a large blow to nuclear energy.
10	EXACT	En marzo de 2011 la central atómica se vio afectada por el terremoto de 9 grados que sacudió y devastó al país, para inmediatamente sufrir el tsunami que arrasó con el sistema de refrigeración de los reactores.	In March 2011, the nuclear plant was damaged by a 9.0 earthquake that shook and devastated the country. It then immediately was pummeled by a tsunami that destroyed the cooling system for the reactors.
11	FUZZY	Esto provocó un recalentamiento que trajo innumerables problemas.	This caused overheating ~~that~~which brought about countless problems.
12		Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar.	
13		El mes pasado el gobierno decidió encarar el problema por su cuenta, desplazando a la operadora privada de la central, TEPCO's, que no encontraba una solución.	
14	EXACT	Para ello estableció una serie de medidas tendientes a frenar las filtraciones.	To do so, a series of measures were established that are designed to slow the leaks.
15	FUZZY	Desplegó un programa que incluye un sofisticado sistema de congelamiento del terreno alrededor de la planta, con tubos que llevan un potente refrigerante.	It unveiled a program that includes a sophisticated ~~ground cooling~~ system for cooling the earth around the plant with pipes carrying a ~~powerful~~ refrigerant.
16	FUZZY	También piensa establecer un proceso novedoso de descontaminación del agua radiactiva, que se viene almacenando peligrosamente en contenedores.	It also aims to establish a new radioactive decontamination process for water that is precariously stored in containers.
17	EXACT	Estos trabajos comenzarán inmediatamente y se espera que concluyan en dos años.	These projects will begin immediately and should be completed in two years.
18		"Hemos realizado monitoreos en la región costera de Fukushima.	

19	FUZZY	Las normas de seguridad de Japón tanto para alimentos como para aguas tienen el nivel más estricto del mundo”.	Japan’s ~~safety~~ security standards for both food and water are the most stringent in the world~~.~~”.
20		Abe sostuvo que “no ha habido problemas relacionados con la salud hasta ahora.	
21	EXACT	No los hay en el presente y no los habrá en el futuro, lo afirmo de la forma más enfática”.	There aren’t any at the moment, and there won’t be any in the future. I cannot emphasize this enough.

Version B

1	FUZZY	En su breve paso por Buenos Aires, el primer ministro japonés Shinzo Abe se encargó de remarcar que su país, y particularmente Tokio, es “seguro” para los Juegos Olímpicos 2020 que acaban de ganar.	During a brief stop~~ in Buenos Aires~~, Shinzo Abe, prime minister of Japan, ~~stressed~~ said that his country, and in particular Tokyo, is "safe" for the 20~~20~~02 Olympic Games that were recently awarded to the city.
2	EXACT	Exultante por haber obtenido lo que vino a buscar, el mandatario aseguró que la crisis desatada en la planta de energía nuclear de Fukushima está absolutamente controlada.	The head of state, thrilled for having achieving what he came for, emphasized that the crisis unfolding at the Fukushima nuclear energy plant is completely under control.
3		También, defendió los millonarios programas oficiales para detener las filtraciones y descontaminar el agua radiactiva.	
4	EXACT	Sostuvo que “el gobierno ha establecido una serie de medidas para reducir la dependencia de la energía nuclear ” en los próximos tres años, al tiempo que aplicarán “los más severos programas científicos de seguridad ” sobre las centrales atómicas.	He maintained that "the government has established a series of measures to reduce the dependency on nuclear energy" over the next three years, while at the same time implementing "the most rigorous scientific programs on safety" concerning nuclear plants.
5	FUZZY	En Japón, la tercera economía mundial, la provisión de energía nuclear es sólo una parte mínima	In Japan, the world's ~~third~~ largest economy, nuclear energy production is only a small part used

		de la que utiliza su potente desarrollo.	by its strong development.
6		Es más, de sus 54 centrales nucleares sólo dos están en funcionamiento.	
7	EXACT	Las otras se encuentran paradas.	The others are shut down.
8		El país se está abasteciendo de otro tipo de energía y promoviendo la conservación y la adopción de energías renovables.	
9	FUZZY	El accidente en Fukushima le dio un golpe mortal a la energía nuclar.	The Fukushima ~~accident~~ disaster dealt a large blow to nuclear energy.
10	FUZZY	En marzo de 2011 la central atómica se vio afectada por el terremoto de 9 grados que sacudió y devastó al país, para inmediatamente sufrir el tsunami que arrasó con el sistema de refrigeración de los reactores.	In March 2011, the nuclear plant was damaged by a 9.0 earthquake that shook and devastated the country~~. It~~ , which was then immediately ~~was~~ pummeled by a tsunami that destroyed the ~~cooling~~ refrigeration system for the reactors.
11		Esto provocó un recalentamiento que trajo innumerables problemas.	
12	EXACT	Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar.	Since then, there have been small leaks of contaminated water that have wound up in the ocean.
13	EXACT	El mes pasado el gobierno decidió encarar el problema por su cuenta, desplazando a la operadora privada de la central, TEPCO’s, que no encontraba una solución.	This past month, the government decided to tackle the problem on its own, replacing the private operator of the plant, TEPCO, which was unable to find a solution.
14	FUZZY	Para ello estableció una serie de medidas tendientes a frenar las filtraciones.	To do so, a series of measures were established that are designed to slow the ~~leaks~~ filters.
15		Desplegó un programa que incluye un sofisticado sistema de congelamiento del terreno alrededor de la planta, con tubos que llevan un potente refrigerante.	
16		También piensa establecer un proceso novedoso de descontaminación del agua radiactiva, que se viene almacenando peligrosamente en contenedores.	

17	FUZZY	Estos trabajos comenzarán inmediatamente y se espera que concluyan en dos años.	These ~~projects~~works will begin immediately and ~~should~~ will be completed in two years.
18	EXACT	“Hemos realizado monitoreos en la región costera de Fukushima.	“We have monitored the coastal region of Fukushima.
19		Las normas de seguridad de Japón tanto para alimentos como para aguas tienen el nivel más estricto del mundo”.	
20	EXACT	Abe sostuvo que “no ha habido problemas relacionados con la salud hasta ahora.	Abe stated that “there have not been any health-related problems up to this point.
21	FUZZY	No los hay en el presente y no los habrá en el futuro, lo afirmo de la forma más enfática”.	There aren’t any at the moment, and there won’t be any~~ in the future~~. I cannot ~~emphasize~~underplay this enough.

Version C

1		En su breve paso por Buenos Aires, el primer ministro japonés Shinzo Abe se encargó de remarcar que su país, y particularmente Tokio, es “seguro” para los Juegos Olímpicos 2020 que acaban de ganar.	
2	FUZZY	Exultante por haber obtenido lo que vino a buscar, el mandatario aseguró que la crisis desatada en la planta de energía nuclear de Fukushima está absolutamente controlada.	The head of state, thrilled for having achieving what came for, emphasized that the crisis ~~unfolding~~ at the Fukushima nuclear energy plant is ~~completely~~ under control.
3	EXACT	También, defendió los millonarios programas oficiales para detener las filtraciones y descontaminar el agua radiactiva.	He also defended the millions of official programs to stop leaks and decontaminate radioactive water.
4	FUZZY	Sostuvo que “el gobierno ha establecido una serie de medidas para reducir la dependencia de la energía nuclear ” en los próximos tres años, al tiempo que aplicarán “los más severos programas científicos de seguridad ” sobre las	He maintained that "the government has established a series of measures to reduce the ~~dependency on~~ need for nuclear energy" over the next three years, ~~while~~ at a time when ~~the same time~~ implementing "the most rigorous

		centrales atómicas.	scientific programs on safety" concerning nuclear plants.
5		En Japón, la tercera economía mundial, la provisión de energía nuclear es sólo una parte mínima de la que utiliza su potente desarrollo.	
6	EXACT	Es más, de sus 54 centrales nucleares sólo dos están en funcionamiento.	In fact, only two of the 54 nuclear power plants are currently operational.
7	FUZZY	Las otras se encuentran paradas.	The others are ~~shut down~~ stopped.
8	EXACT	El país se está abasteciendo de otro tipo de energía y promoviendo la conservación y la adopción de energías renovables.	The country is supplied by another type of energy and is promoting energy conservation and the adoption of renewable energy.
9		El accidente en Fukushima le dio un golpe mortal a la energía nuclar.	
10		En marzo de 2011 la central atómica se vio afectada por el terremoto de 9 grados que sacudió y devastó al país, para inmediatamente sufrir el tsunami que arrasó con el sistema de refrigeración de los reactores.	
11	EXACT	Esto provocó un recalentamiento que trajo innumerables problemas.	This caused overheating that brought about countless problems.
12	FUZZY	Desde ese momento hay pequeñas filtraciones de agua contaminada que terminan en el mar.	Since then, there have been ~~small~~ leaks of contaminated water that have ~~wound up~~ ended in the ~~ocean~~ sea.
13	FUZZY	El mes pasado el gobierno decidió encarar el problema por su cuenta, desplazando a la operadora privada de la central, TEPCO's, que no encontraba una solución.	This past month, the government decided to tackle the problem ~~on its own~~, ~~replacing~~ removing the private operator of the plant, TEPCO, which was unable to find an answer ~~solution~~.
14		Para ello estableció una serie de medidas tendientes a frenar las filtraciones.	
15	EXACT	Desplegó un programa que incluye un sofisticado sistema de congelamiento del terreno alrededor de la planta, con tubos que llevan un potente refrigerante.	It unveiled a program that includes a sophisticated ground cooling system around the plant with pipes carrying a powerful refrigerant.

16	EXACT	También piensa establecer un proceso novedoso de descontaminación del agua radiactiva, que se viene almacenando peligrosamente en contenedores.	It also aims to establish a new radioactive water decontamination process that is precariously stored in containers.
17		Estos trabajos comenzarán inmediatamente y se espera que concluyan en dos años.	
18	FUZZY	“Hemos realizado monitoreos en la región costera de Fukushima.	“We have ~~monitored~~ reviewed the coastal region of Fukushima.
19	EXACT	Las normas de seguridad de Japón tanto para alimentos como para aguas tienen el nivel más estricto del mundo”.	Japan’s safety standards for both food and water are the most stringent in the world.”
20	FUZZY	Abe sostuvo que “no ha habido problemas relacionados con la salud hasta ahora.	Abe ~~stated~~ claimed that “there have not been any health-related problems up to this point.
21		No los hay en el presente y no los habrá en el futuro, lo afirmo de la forma más enfática”.	

Appendix F:

Experimental instrument – Version setup

Version A

1	EXACT
2	
3	FUZZY
4	
5	EXACT
6	FUZZY
7	
8	FUZZY
9	EXACT
10	EXACT
11	FUZZY
12	
13	
14	EXACT
15	FUZZY
16	FUZZY
17	EXACT
18	
19	FUZZY
20	
21	EXACT

Version B

1	FUZZY
2	EXACT
3	
4	EXACT
5	FUZZY
6	
7	EXACT

8	
9	FUZZY
10	FUZZY
11	
12	EXACT
13	EXACT
14	FUZZY
15	
16	
17	FUZZY
18	EXACT
19	
20	EXACT
21	FUZZY

Version C

1	
2	FUZZY
3	EXACT
4	FUZZY
5	
6	EXACT
7	FUZZY
8	EXACT
9	
10	
11	EXACT
12	FUZZY
13	FUZZY
14	
15	EXACT
16	EXACT
17	
18	FUZZY
19	EXACT
20	FUZZY
21	

Appendix G:

Experiment Recruitment Script

Hello,

Based on your survey responses, you are eligible to participate in the second part of a research study that will examine effort in translation when using computer-assisted translation software. This study is being conducted by Dr. Keiran Dunne, principal investigator, and co-investigator, Ph.D. candidate Christopher Mellinger, of Kent State University, and it has been approved by the Kent State University Institutional Review Board.

The second part of the study consists of translating a text of approximately 400 words from Spanish to English using a web-based computer-assisted translation tool and without using any external resources. This translation should take no longer than 90 minutes to complete. Your participation is entirely voluntary. You can withdraw from the study at any time. No risks are anticipated beyond those encountered in everyday life. Data gathered during the study will be shared with no one.

To take part in the second part of the study, please contact me at cmellin2@kent.edu to receive a unique log-in name and password to complete the translation. If you are no longer interested in taking part in this research study, please let me know and your name will be removed from the list.

Please feel free to contact me if you have any questions. Thank you very much in advance.

Best regards,

Chris Mellinger
cmellin2@kent.edu

Computer-Assisted Translation: An Empirical Investigation of Cognitive Effort

Welcome to "Effort in Translation," a web-based experiment designed to better understand how people translate when using computer-assisted translation (CAT) tools. Before taking part in this study, please read the consent form below and click on the "I Agree" button at the bottom of the page if you understand the statements and freely consent to participate in the study.

Consent Form

This study is being conducted by Ph.D. candidate Christopher Mellinger and Dr. Keiran Dunne of Kent State University, and it has been approved by the Kent State University Institutional Review Board. No deception is involved, and the study involves no more than minimal risk to participants (i.e., the level of risk encountered in daily life).

Participation in the study typically takes no more than 90 minutes and is strictly anonymous. Participants will be asked to translate a text of approximately 400 words from Spanish to English without using any resources other than those provided by the computer-assisted translation system. The text will appear on the left-hand side of the screen and you will have the option of either translating the segment or editing the suggested translation that appears in the box on the right-hand side of the screen.

All responses will be treated with absolute confidentiality, and in no case will responses from individual participants be identified. Rather, all data will be pooled and published in aggregate form only. Any reference to a specific participant will be made using an anonymous name. Participants should be aware, however, that the experiment is not being run from a "secure" https server of the kind typically used to handle credit card transactions, so there is a small possibility that responses could be viewed by unauthorized third parties (e.g., computer hackers).

Participation in this study will not entail any risk to you beyond those risks encountered in everyday life, and no adverse reactions have been reported. Participation is voluntary and refusal to take part in the study involves no penalty or loss of benefits to which you are otherwise entitled. You may withdraw from the study at any time.

If you have further questions about this study or your rights, or if you wish to lodge a complaint or concern, you may contact the principal investigator, Dr. Keiran Dunne, or the co-investigator, Ph.D. candidate Christopher Mellinger, at (908) 399-9252; or the Kent State University Institutional Review Board, at (330) 672-2704.

If you are 18 years of age or older, understand the statements above, and freely consent to participate in the study, click on the "I Agree" button to begin the experiment.

I Agree   I Do Not Agree

Appendix I: Summary of Means (SD) and one-way, repeated measures ANOVA statistics

	Average Pause Ratio	Keystrokes	Mouseclicks	Total Pause Time (sec)	Total Segment Time (sec)
Exact	9.66 (18.68)	77.76 (206.85)	3.15 (4.43)	85.89 (91.38)	77.45 (98.42)
Fuzzy	6.95 (20.04)	90.56 (198.35)	2.93 (3.35)	99.63 (137.96)	82.80 (101.49)
New	1.23 (.605)	230.27 (151.64)	3.00 (3.61)	163.92 (176.37)	153.05 (154.54)
ANOVA[a]	$F(1.55, 96.185) = 4.556$, $p = .02$[b]	$F(1.306, 80.955) = 17.766$, $p < .001$[b,c]	$F(1.685, 104.496) = .089$, $p=.884$	$F(1.529, 94.784) = 5.828$, $p=.008$[b]	$F(1.468, 91.027) = 8.032$, $p=.002$[b,c]

[a] Greenhouse-Geisser correction
[b] Difference between Exact and New matches
[c] Difference between Fuzzy and New matches

Average Pause Ratio

Lacruz, Shreve, and Angelone (2012) propose the average pause ratio as a measure of the cognitive effort exerted by translators. The authors establish this metric as an alternative to the temporal measure of the total amount of time spent in pauses, since “pauses are of variable length, and a large number of short pauses will likely indicate a different cognitive processing/effort pattern than a single pause of the same overall duration” (2012: n.p.). The researchers’ rationale suggests that the total amount of pause time does not sufficiently take into account the overall length of the segment, nor does it differentiate between clusters of pauses and a single longer pause. The formula to calculate the average pause ratio in each segment is as follows:

$$\text{Average Pause Ratio (APR)} = \frac{\text{Average time per pause}}{\text{Average time per word}}$$

As can be seen in this formula, the number of pauses is taken into account by averaging the amount of time per pause. Segment length is also addressed by averaging the amount of time spent per word in the segment. When the average amount of time per pause is compared to the average time per word in the segment, the resulting ratio is the average pause ratio. Higher values of this ratio indicate less effortful segments, while lower values suggest the exertion of more effort.

Correlations

Notes		
Output Created		14-FEB-2014 16:10:52
Comments		
Input	Active Dataset	DataSet1
	Filter	
	Weight	
	Split File	Participant
	N of Rows in Working Data File	189
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
Syntax		CORRELATIONS /VARIABLES=Rating Keystrokes Mouseclicks PauseRatio PauseTime SegmentTime /PRINT=TWOTAIL NOSIG /STATISTICS DESCRIPTIVES /MISSING=PAIRWISE.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.07

[DataSet1]

Participant = 1.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	4.5714	.67612	21
Keystrokes	92.4762	122.65913	21
Mouseclicks	2.1905	3.44411	21
PauseRatio	4.8850	5.26219	21
PauseTime	166.7406	217.23196	21
SegmentTime	174.4354	219.75546	21

a. Participant = 1.00

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.120	-.865**	.207	-.154	-.150
	Sig. (2-tailed)	.603	.000	.367	.504	.516	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.120	1	.304	-.553**	.660**	.728**

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
	Sig. (2-tailed)	.603		.181	.009	.001	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.865**	.304	1	-.210	.280	.289
	Sig. (2-tailed)	.000	.181		.362	.219	.204
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.207	-.553**	-.210	1	-.350	-.382
	Sig. (2-tailed)	.367	.009	.362		.120	.088
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.154	.660**	.280	-.350	1	.995**
	Sig. (2-tailed)	.504	.001	.219	.120		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.150	.728**	.289	-.382	.995**	1
	Sig. (2-tailed)	.516	.000	.204	.088	.000	
	N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 1.00

Participant = 2.00

Descriptive Statistics[a]			
	Mean	Std. Deviation	N
Rating	3.0952	1.17918	21
Keystrokes	79.7619	76.08936	21
Mouseclicks	3.1429	2.95442	21
PauseRatio	3.4572	2.89486	21
PauseTime	58.8542	50.70303	21
SegmentTime	57.2369	39.02288	21

a. Participant = 2.00

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.746**	-.105	.767**	-.214	-.347
	Sig. (2-tailed)	.000	.652	.000	.351	.123	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.746**	1	.256	-.541*	.356	.720**
	Sig. (2-tailed)	.000		.263	.011	.113	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.105	.256	1	-.233	.728**	.659**
	Sig. (2-tailed)	.652	.263		.309	.000	.001
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.767**	-.541*	-.233	1	-.204	-.383
	Sig. (2-tailed)	.000	.011	.309		.375	.087
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.214	.356	.728**	-.204	1	.726**
	Sig. (2-tailed)	.351	.113	.000	.375		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.347	.720**	.659**	-.383	.726**	1
	Sig. (2-tailed)	.123	.000	.001	.087	.000	

	Sig. (2-tailed)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
	N	21	21	21	21	21	21
**. Correlation is significant at the 0.01 level (2-tailed).							
*. Correlation is significant at the 0.05 level (2-tailed).							
a. Participant = 2.00							

Participant = 3.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	4.1429	.47809	21
Keystrokes	99.3333	118.77724	21
Mouseclicks	5.8571	5.12138	21
PauseRatio	2.5474	1.86828	21
PauseTime	157.3292	145.21738	21
SegmentTime	105.9356	101.66655	21
a. Participant = 3.00			

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.195	-.379	.275	-.255	-.267
	Sig. (2-tailed)	.396	.090	.229	.264	.241	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.195	1	.701**	-.386	.673**	.761**
	Sig. (2-tailed)	.396		.000	.084	.001	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.379	.701**	1	-.282	.855**	.863**
	Sig. (2-tailed)	.090	.000		.216	.000	.000
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.275	-.386	-.282	1	-.246	-.365
	Sig. (2-tailed)	.229	.084	.216		.283	.104
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.255	.673**	.855**	-.246	1	.964**
	Sig. (2-tailed)	.264	.001	.000	.283		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.267	.761**	.863**	-.365	.964**	1
	Sig. (2-tailed)	.241	.000	.000	.104	.000	
	N	21	21	21	21	21	21
**. Correlation is significant at the 0.01 level (2-tailed).							
a. Participant = 3.00							

Participant = 4.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N

Rating	4.7619	.43644	21
Keystrokes	125.6667	175.05009	21
Mouseclicks	2.2381	1.97243	21
PauseRatio	11.4516	34.43161	21
PauseTime	105.5571	111.95257	21
SegmentTime	104.4960	105.71120	21
a. Participant = 4.00			

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.584**	-.279	.171	-.634**	-.674**
	Sig. (2-tailed)	.005	.220	.460	.002	.001	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.584**	1	.719**	-.227	.721**	.787**
	Sig. (2-tailed)	.005		.000	.321	.000	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.279	.719**	1	-.311	.757**	.771**
	Sig. (2-tailed)	.220	.000		.170	.000	.000
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.171	-.227	-.311	1	-.251	-.285
	Sig. (2-tailed)	.460	.321	.170		.272	.211
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.634**	.721**	.757**	-.251	1	.979**
	Sig. (2-tailed)	.002	.000	.000	.272		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.674**	.787**	.771**	-.285	.979**	1
	Sig. (2-tailed)	.001	.000	.000	.211	.000	
	N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 4.00

Participant = 5.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	4.0000	.83666	21
Keystrokes	393.4762	382.41778	21
Mouseclicks	7.5238	6.05491	21
PauseRatio	1.6266	2.09984	21
PauseTime	162.9777	138.88751	21
SegmentTime	152.2335	142.10032	21

a. Participant = 5.00

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.528*	-.454*	.327	-.381	-.510*
	Sig. (2-tailed)	.014	.039	.147	.088	.018	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.528*	1	.689**	-.402	.784**	.938**
	Sig. (2-tailed)	.014		.001	.071	.000	.000

	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.454*	.689**	1	-.346	.722**	.749**
	Sig. (2-tailed)	.039	.001		.125	.000	.000
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.327	-.402	-.346	1	-.325	-.379
	Sig. (2-tailed)	.147	.071	.125		.150	.090
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.381	.784**	.722**	-.325	1	.925**
	Sig. (2-tailed)	.088	.000	.000	.150		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.510*	.938**	.749**	-.379	.925**	1
	Sig. (2-tailed)	.018	.000	.000	.090	.000	
	N	21	21	21	21	21	21
*. Correlation is significant at the 0.05 level (2-tailed).							
**. Correlation is significant at the 0.01 level (2-tailed).							
a. Participant = 5.00							

Participant = 6.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	4.2857	.78376	21
Keystrokes	100.1429	190.82303	21
Mouseclicks	1.4762	1.50396	21
PauseRatio	10.4687	22.20278	21
PauseTime	136.9468	187.66701	21
SegmentTime	103.9429	147.60504	21
a. Participant = 6.00			

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	.502*	.515*	.110	.441*	.409
	Sig. (2-tailed)	.021	.017	.636	.045	.066	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	.502*	1	.821**	-.230	.932**	.973**
	Sig. (2-tailed)	.021		.000	.315	.000	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	.515*	.821**	1	-.174	.817**	.823**
	Sig. (2-tailed)	.017	.000		.451	.000	.000
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.110	-.230	-.174	1	-.149	-.249
	Sig. (2-tailed)	.636	.315	.451		.518	.276
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	.441*	.932**	.817**	-.149	1	.940**
	Sig. (2-tailed)	.045	.000	.000	.518		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	.409	.973**	.823**	-.249	.940**	1
	Sig. (2-tailed)						

		.066	.000	.000	.276	.000	
	N	21	21	21	21	21	21
*. Correlation is significant at the 0.05 level (2-tailed).							
**. Correlation is significant at the 0.01 level (2-tailed).							
a. Participant = 6.00							

Participant = 7.00

Descriptive Statistics[a]			
	Mean	Std. Deviation	N
Rating	3.5714	1.24786	21
Keystrokes	122.7619	105.25536	21
Mouseclicks	1.6667	1.42595	21
PauseRatio	5.7416	9.71470	21
PauseTime	100.3269	80.73519	21
SegmentTime	97.0629	74.32425	21
a. Participant = 7.00			

Correlations[a]							
		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.814**	-.028	.559**	-.595**	-.669**
	Sig. (2-tailed)	.000	.904	.008	.004	.001	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.814**	1	.232	-.546*	.809**	.872**
	Sig. (2-tailed)	.000		.311	.010	.000	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.028	.232	1	-.080	.159	.160
	Sig. (2-tailed)	.904	.311		.729	.490	.489
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.559**	-.546*	-.080	1	-.414	-.482*
	Sig. (2-tailed)	.008	.010	.729		.062	.027
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.595**	.809**	.159	-.414	1	.958**
	Sig. (2-tailed)	.004	.000	.490	.062		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.669**	.872**	.160	-.482*	.958**	1
	Sig. (2-tailed)	.001	.000	.489	.027	.000	
	N	21	21	21	21	21	21
**. Correlation is significant at the 0.01 level (2-tailed).							
*. Correlation is significant at the 0.05 level (2-tailed).							
a. Participant = 7.00							

Participant = 8.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	3.7619	.76842	21
Keystrokes	88.7143	107.80637	21
Mouseclicks	2.1429	1.62129	21
PauseRatio	8.8496	22.65410	21
PauseTime	71.5976	89.12761	21
SegmentTime	60.6007	54.73069	21
a. Participant = 8.00			

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.081	-.212	.405	-.451*	-.391
	Sig. (2-tailed)	.727	.356	.068	.040	.080	
	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.081	1	.177	-.290	.357	.744**
	Sig. (2-tailed)	.727		.442	.201	.112	.000
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.212	.177	1	-.070	.276	.324
	Sig. (2-tailed)	.356	.442		.764	.225	.153
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.405	-.290	-.070	1	-.199	-.325
	Sig. (2-tailed)	.068	.201	.764		.387	.151
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.451*	.357	.276	-.199	1	.823**
	Sig. (2-tailed)	.040	.112	.225	.387		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.391	.744**	.324	-.325	.823**	1
	Sig. (2-tailed)	.080	.000	.153	.151	.000	
	N	21	21	21	21	21	21

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
a. Participant = 8.00

Participant = 9.00

Descriptive Statistics[a]

	Mean	Std. Deviation	N
Rating	4.2857	.56061	21
Keystrokes	93.4286	133.79969	21
Mouseclicks	1.0476	1.24403	21
PauseRatio	4.5142	3.26271	21
PauseTime	88.0363	161.04103	21
SegmentTime	83.9543	113.94590	21
a. Participant = 9.00			

Correlations[a]

		Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Rating	Pearson Correlation	1	-.126	-.307	.499*	-.172	-.203
	Sig. (2-tailed)	.585	.175	.021	.455	.378	
	N	21	21	21	21	21	21

	N	21	21	21	21	21	21
Keystrokes	Pearson Correlation	-.126	1	.157	-.579**	.156	.379
	Sig. (2-tailed)	.585		.498	.006	.501	.090
	N	21	21	21	21	21	21
Mouseclicks	Pearson Correlation	-.307	.157	1	-.248	.299	.358
	Sig. (2-tailed)	.175	.498		.279	.188	.111
	N	21	21	21	21	21	21
PauseRatio	Pearson Correlation	.499*	-.579**	-.248	1	-.201	-.323
	Sig. (2-tailed)	.021	.006	.279		.383	.153
	N	21	21	21	21	21	21
PauseTime	Pearson Correlation	-.172	.156	.299	-.201	1	.954**
	Sig. (2-tailed)	.455	.501	.188	.383		.000
	N	21	21	21	21	21	21
SegmentTime	Pearson Correlation	-.203	.379	.358	-.323	.954**	1
	Sig. (2-tailed)	.378	.090	.111	.153	.000	
	N	21	21	21	21	21	21
*. Correlation is significant at the 0.05 level (2-tailed).							
**. Correlation is significant at the 0.01 level (2-tailed).							
a. Participant = 9.00							

NONPAR CORR
/VARIABLES=Rating Keystrokes Mouseclicks PauseRatio PauseTime SegmentTime
/PRINT=SPEARMAN TWOTAIL NOSIG
/MISSING=PAIRWISE.

Nonparametric Correlations

Notes		
Output Created		14-FEB-2014 16:10:53
Comments		
Input	Active Dataset	DataSet1
	Filter	
	Weight	
	Split File	Participant
	N of Rows in Working Data File	189
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
Syntax		NONPAR CORR /VARIABLES=Rating Keystrokes Mouseclicks PauseRatio PauseTime SegmentTime /PRINT=SPEARMAN TWOTAIL NOSIG /MISSING=PAIRWISE.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.02
	Number of Cases Allowed	87381 cases[a]

a. Based on availability of workspace memory

Participant = 1.00

Correlations[a]								
			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.079	-.862**	.056	-.466*	-.461*
		Sig. (2-tailed)	.	.732	.000	.809	.033	.035
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.079	1.000	.014	-.848**	.656**	.690**
		Sig. (2-tailed)	.732	.	.953	.000	.001	.001
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.862**	.014	1.000	.053	.431	.399
		Sig. (2-tailed)	.000	.953	.	.819	.051	.073
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.056	-.848**	.053	1.000	-.447*	-.491*
		Sig. (2-tailed)	.809	.000	.819	.	.042	.024
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.466*	.656**	.431	-.447*	1.000	.981**
		Sig. (2-tailed)	.033	.001	.051	.042	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.461*	.690**	.399	-.491*	.981**	1.000
		Sig. (2-tailed)	.035	.001	.073	.024	.000	.
		N	21	21	21	21	21	21
**. Correlation is significant at the 0.01 level (2-tailed).								
*. Correlation is significant at the 0.05 level (2-tailed).								
a. Participant = 1.00								

Participant = 2.00

Correlations[a]								
			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.816**	-.128	.810**	-.222	-.444*
		Sig. (2-tailed)	.	.000	.580	.000	.334	.044
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.816**	1.000	.387	-.656**	.483*	.793**
		Sig. (2-tailed)	.000	.	.083	.001	.026	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.128	.387	1.000	.031	.599**	.501*
		Sig. (2-tailed)	.580	.083	.	.894	.004	.021
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.810**	-.656**	.031	1.000	-.070	-.397
		Sig. (2-tailed)	.000	.001	.894	.	.763	.074
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.222	.483*	.599**	-.070	1.000	.738**
		Sig. (2-tailed)	.334	.026	.004	.763	.	.000

		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.444*	.793**	.501*	-.397	.738**	1.000
		Sig. (2-tailed)	.044	.000	.021	.074	.000	.
		N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. Participant = 2.00

Participant = 3.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.282	-.517*	.187	-.382	-.444*
		Sig. (2-tailed)	.	.216	.016	.417	.088	.044
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.282	1.000	.671**	-.477*	.566**	.697**
		Sig. (2-tailed)	.216	.	.001	.029	.007	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.517*	.671**	1.000	-.378	.791**	.914**
		Sig. (2-tailed)	.016	.001	.	.091	.000	.000
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.187	-.477*	-.378	1.000	-.356	-.484*
		Sig. (2-tailed)	.417	.029	.091	.	.113	.026
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.382	.566**	.791**	-.356	1.000	.930**
		Sig. (2-tailed)	.088	.007	.000	.113	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.444*	.697**	.914**	-.484*	.930**	1.000
		Sig. (2-tailed)	.044	.000	.000	.026	.000	.
		N	21	21	21	21	21	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 3.00

Participant = 4.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.611**	-.246	.443*	-.517*	-.628**
		Sig. (2-tailed)	.	.003	.282	.044	.016	.002
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.611**	1.000	.685**	-.870**	.750**	.858**
		Sig. (2-tailed)	.003	.	.001	.000	.000	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.246	.685**	1.000	-.541*	.747**	.769**
		Sig. (2-tailed)	.282	.001	.	.011	.000	.000

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.443*	-.870**	-.541*	1.000	-.619**	-.816**
		Sig. (2-tailed)	.044	.000	.011	.	.003	.000
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.517*	.750**	.747**	-.619**	1.000	.903**
		Sig. (2-tailed)	.016	.000	.000	.003	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.628**	.858**	.769**	-.816**	.903**	1.000
		Sig. (2-tailed)	.002	.000	.000	.000	.000	.
		N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
a. Participant = 4.00

Participant = 5.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.462*	-.461*	.424	-.356	-.405
		Sig. (2-tailed)	.	.035	.036	.056	.113	.069
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.462*	1.000	.687**	-.603**	.752**	.859**
		Sig. (2-tailed)	.035	.	.001	.004	.000	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.461*	.687**	1.000	-.509*	.731**	.754**
		Sig. (2-tailed)	.036	.001	.	.018	.000	.000
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.424	-.603**	-.509*	1.000	-.332	-.513*
		Sig. (2-tailed)	.056	.004	.018	.	.141	.017
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.356	.752**	.731**	-.332	1.000	.927**
		Sig. (2-tailed)	.113	.000	.000	.141	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.405	.859**	.754**	-.513*	.927**	1.000
		Sig. (2-tailed)	.069	.000	.000	.017	.000	.
		N	21	21	21	21	21	21

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
a. Participant = 5.00

Participant = 6.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	.649**	.587**	-.468*	.413	.428
		Sig. (2-tailed)	.	.001	.005	.032	.063	.053
		N	21	21	21	21	21	21

		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	.649**	1.000	.714**	-.838**	.648**	.776**
		Sig. (2-tailed)	.001	.	.000	.000	.001	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	.587**	.714**	1.000	-.533*	.318	.453*
		Sig. (2-tailed)	.005	.000	.	.013	.160	.039
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	-.468*	-.838**	-.533*	1.000	-.397	-.692**
		Sig. (2-tailed)	.032	.000	.013	.	.074	.001
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	.413	.648**	.318	-.397	1.000	.756**
		Sig. (2-tailed)	.063	.001	.160	.074	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	.428	.776**	.453*	-.692**	.756**	1.000
		Sig. (2-tailed)	.053	.000	.039	.001	.000	.
		N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

a. Participant = 6.00

Participant = 7.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.820**	-.053	.935**	-.636**	-.751**
		Sig. (2-tailed)	.	.000	.821	.000	.002	.000
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.820**	1.000	.172	-.909**	.805**	.902**
		Sig. (2-tailed)	.000	.	.456	.000	.000	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.053	.172	1.000	-.142	.293	.293
		Sig. (2-tailed)	.821	.456	.	.538	.198	.198
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.935**	-.909**	-.142	1.000	-.694**	-.804**
		Sig. (2-tailed)	.000	.000	.538	.	.000	.000
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.636**	.805**	.293	-.694**	1.000	.936**
		Sig. (2-tailed)	.002	.000	.198	.000	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.751**	.902**	.293	-.804**	.936**	1.000
		Sig. (2-tailed)	.000	.000	.198	.000	.000	.
		N	21	21	21	21	21	21

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 7.00

Participant = 8.00

Correlations[a]

Correlations

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.305	-.244	.317	-.433*	-.385
		Sig. (2-tailed)	.	.179	.287	.161	.050	.085
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.305	1.000	.379	-.885**	.736**	.807**
		Sig. (2-tailed)	.179	.	.090	.000	.000	.000
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.244	.379	1.000	-.265	.426	.488*
		Sig. (2-tailed)	.287	.090	.	.246	.054	.025
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.317	-.885**	-.265	1.000	-.621**	-.699**
		Sig. (2-tailed)	.161	.000	.246	.	.003	.000
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.433*	.736**	.426	-.621**	1.000	.931**
		Sig. (2-tailed)	.050	.000	.054	.003	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.385	.807**	.488*	-.699**	.931**	1.000
		Sig. (2-tailed)	.085	.000	.025	.000	.000	.
		N	21	21	21	21	21	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 8.00

Participant = 9.00

Correlations[a]

			Rating	Keystrokes	Mouseclicks	PauseRatio	PauseTime	SegmentTime
Spearman's rho	Rating	Correlation Coefficient	1.000	-.325	-.415	.437*	-.291	-.285
		Sig. (2-tailed)	.	.150	.061	.048	.201	.211
		N	21	21	21	21	21	21
	Keystrokes	Correlation Coefficient	-.325	1.000	.195	-.900**	.432	.635**
		Sig. (2-tailed)	.150	.	.398	.000	.051	.002
		N	21	21	21	21	21	21
	Mouseclicks	Correlation Coefficient	-.415	.195	1.000	-.082	.228	.251
		Sig. (2-tailed)	.061	.398	.	.722	.321	.272
		N	21	21	21	21	21	21
	PauseRatio	Correlation Coefficient	.437*	-.900**	-.082	1.000	-.235	-.447*
		Sig. (2-tailed)	.048	.000	.722	.	.305	.042
		N	21	21	21	21	21	21
	PauseTime	Correlation Coefficient	-.291	.432	.228	-.235	1.000	.958**
		Sig. (2-tailed)	.201	.051	.321	.305	.	.000
		N	21	21	21	21	21	21
	SegmentTime	Correlation Coefficient	-.285	.635**	.251	-.447*	.958**	1.000
		Sig. (2-tailed)	.211	.002	.272	.042	.000	.
		N	21	21	21	21	21	21

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Participant = 9.00