City University of Hong Kong

An Eye Tracking Study of Cognitive Effort Allocation Across Translation Subtasks

Project Report

Submitted to
Department of Linguistics and Translation

for
The Course LT4235 Project

by
WU Ting-Wei, Tiffany

Supervised by
Dr. Kit Chunyu

May 2016

Abstract

As new technologies have been transforming the landscape of translation industry, translation has become a particular field of interest for research aiming to improve its workflow and optimize its working environment. The objectives of this project are to identify distribution patterns of translators' cognitive effort to different subtasks of translation during the translation process with the aid of a video-based eye-tracking system. This project report first reviews a number of models used in previous researches, as well as the mechanism and interpretation of eye tracking systems and data. The empirical investigation in this project attempts three major analyses: (1) analysis of generalized patterns of attentional distribution to compare the levels of cognitive effort required in respective subtasks; (2) analysis of the fixation duration and pupil size to investigate the cognitive workload among different subtasks under investigation and also the relationship between key events and fixation durations; (3) analysis of translators' working styles of with respect to cross interest area saccades and duration of follow-up fixations. The analysis results show that most attentional shifts take place between ST and TT, while no significantly different patterns of dictionary lookup activities prompted by the needs for comprehension versus production are identified. The analyses are presented together with discussions and references to previous models, and possible avenues to future research are also sketched.

Table of Contents

Abstract--2

Table of Content--3

1. Introduction---4

2. Methodology---8

3. Experiment Design---16

4. Data Analysis--22

5. Conclusion--40

6. Appendix--43

References---49

1. Introduction

The majority of translation studies have traditionally focused on texts, linguistics and cultures, whereas less attention has been directed to the cognitive aspects of translators during translating. Jääskeläinen (2000) commented on the situation that "process-oriented research efforts may lack the explanatory power required to draw reliable generalizations which are necessary for building viable theories and creating testable hypotheses". An empirically-grounded cognitive model of translation processes can not only be instrumental for the improvement of translation education, but also provide a basis for the development of translation assistance tools and for more successful interaction between human translators and the technology (Carl et al., 2011).

The application of eye tracking to translation studies has offered quantitative explanations for various questions in researches as well as provided statistical evidence for analysis and classification of the acknowledged "highly individualistic" behaviors (Boehm,1993), such as writing process. Recognizing its potentials, eye tracking is adopted in this project as a major method of data collection, aiming to discover the general distributional patterns and different levels of cognitive resources required across several subtasks during translation process.

This section briefly reviews the development of translation process studies, and discusses several models proposed in previous researches.

1.1 Translation Process in Progress

Since 1950s, a variety of approaches have been implemented to analyze textual elements of translation and characterize the translation process. Translation shift analysis, as proposed by Vinay and Darbelnet (1958) and termed by Catford (1965), aims to describe the phenomenon of translation by analyzing and categorizing the strategies and "shifts" with regard to source text (ST) and target text (TT) pair. Linguistic approaches of this kind describe what constitutes the textual product (Munday, 2001) without considering the cognitive aspects of human translators.

Later research interests in translation process were substantially directed to the cognitive process of translators. As Bell (1991:43) argued, "focus on the description of the process and/or the translator […] form the twin issues which translation theory must address: how the process takes place and what knowledge and skills the translator must possess in order to carry it out." While early models, such as interpretive model and relevance theory, remained rather hypothetical with no empirical evidence, other theorists attempted to gather observational data for the description of the decision-making process of translators. For example, think-aloud protocols, introduced into translation process studies in the mid-1980s, brought the framework from cognitive psychology: a translator was asked to verbalize his/her thoughts while working on a translation task. The validity of data by this modality, nonetheless, has been questioned, since considerable slowing down of processing speed (Jackobsen, 2003) and more alarmingly, changing of the structure or course of translators' processing, such as segmentation, were reported (Jääskeläinen, 2011).

In recent decades, with the emergence and proliferation of new technologies, technological aids such as eye tracking and key-logging have transformed the research and the process of translation. While more efforts should be put into refining the experimental design and establishing theoretical and methodological framework, increasing interests and the potentials of novel methodologies have been recognized, and several models have been proposed, as discussed in the following paragraphs.

1.2 Modelling the Translation Process

Various models and facets of translation process have been proposed so far by attempts in the fields of cognitive psychology and information processing. These models share similar components concerning translation process but vary in the nature of relationship between the components.

Gile (1995) emphasized on the sequentiality of translation and proposed a process model that a translation unit is first read and comprehended as a "meaning hypothesis", which is subsequently reformulated into target language (TL). The two phases, comprehension and reformulation, consist of several small steps. After reformulating the meaning hypothesis into TL, fidelity and acceptability are evaluated and continuously revised until a satisfactory TT version has been arrived at. Gile (1995: 110) also noted, however, that "[o]ftentimes, the translator does not test the meaning hypothesis until after verbalizing it in the target language", suggesting that comprehension and reformulation activities are more integrated than implied by the model.

Contrary to Gile's perspective of sequentiality, Danks & Griffin (1997) argued that full comprehension of ST is not necessary for moving on to producing TT, but instead, the translator may work on various possible solutions while still attempting to comprehend ST. Yet, in line with Gile, the possible solutions will be repeatedly revised and evaluated until a satisfactory output is achieved. Empirical evidence from Jakobsen and Jensen (2008), showing increased gaze activity in reading for translation than reading for normal comprehension, indicates the translator may indeed already engage in translation-related activities during ST comprehension (Dragsted & Carl, 2013).

On the other hand, the monitor model proposed by Tirkkonen-Condit (2005) suggests that conscious decision-making is triggered by a monitor that alerts about the problem in the outcome, otherwise literal translation is adopted as default. Similarly, Kring's (1986) model focuses on translation problems, suggesting that text segments involving a translation problem elicit the application of translation strategies.

Further, based on eye tracking and key-logging technology, Jakobsen (2011) identified the indications of a recurrent "micro-cycle", consisting of six steps, each of which may be skipped or repeated several times. A cycle consists of comprehension of the chunk to be translated (step 1). Then the gaze is shifted to TT to locate the position (step 2) and the translation is typed and monitored (step 3 and 4). And subsequently, the translator shifts back to ST and the current ST chunk is located and reread (step 5 and 6).

Numerous models have been proposed so far, while unanimous and comprehensive descriptions of translation process with solid empirical support have yet been recognized. These models, nonetheless, have pointed to some prospective directions for future researches and brought about fruitful discussions and improvements concerning the experimental design and methodology.

1.3 Structure of the Project Report

The structure of this project report is organized as follows. The following section will address eye tracking in terms of its methodological aspects. Section 3 illustrates the experimental design of this project and identifies some difficulties of eye tracking experiments for examining the translation process. Section 4 presents data analysis in regard of different respects of the translation process. And finally, conclusions concerning the findings and possible directions for future studies will be given in Section 5.

2. Methodology

Eye tracking has provided researchers an aid to understand how human visual systems work and how our minds process visual feedback. It has been applied to various fields including marketing, cognitive psychology, human factors and the broad field of human-computer interaction. However, it is not until the last half-century that the significant technological advancement has allowed us to capture and visualize cognitive processes and to accurately observe visual perception (Bergstrom & Schall, 2014). This section discusses the development of eye-tracking technology and how the current systems, including the one used in this project, work. An overview of data interpretation is provided to facilitate the discussion of data analysis in the later section.

2.1 Development in Eye Tracking Technology

Early attempts at tracking eye movements began in the late 1800s by putting a plaster cup attachment covering over the eyeball so as to capture from one eye the conjugate movements made during reading with the other eye. Early eye-tracking studies were reserved for understanding the basic hypotheses of how the brain and visual system cooperate. In the 1940s, photographic method was adopted to track eye movement. Some earliest eye-tracking usability studies were carried out, such as the study of pilots' eyes movement as they used cockpit controls and instruments. This kind of usability tests sought to study users' interaction with an interface for the purpose of improving interface design (Jacob & Karn, 2004; as cited in Bergstrom & Schall, 2014).

Video-based eye trackers have developed throughout several decades. In the 60s and 70s, the primitive and intrusive apparatus, which required a head restraint and bite bar, still made participants suffer from an uncomfortable environment, and researchers were not able to simulate a comfortable and realistic environment for users. The late 1990s brought about the modern systems of eye tracker that are still used in the industry in recent days. Advancements in both hardware and software made this technology available to not only academic but commercial usage. Easy and rapid calibration and improved accuracy enable an operator with minimal training to carry out experiment sessions. As highly automated software can automatically interpret raw data of eye tracking in terms of various measures, including fixations, saccades, etc., data analysis has become less time-consuming than before. As a consequence, eye tracking has been widely implemented in research and industry, for example, for the development of virtual reality devices as well as for the productivity study of human-computer interactive software such as computer-aided translation tools.

2.2 Mechanism

EyeLink 1000 of SR Research with fixed head mount is used for this project, and Screen Recorder of its software package is adopted for video recordings. This part discusses eye tracking mechanism in general instead of restricting to the specific device for this project so as to present a generic discussion on technical aspects of the technology.

Current commercial eye-tracking systems employ video imaging to determine the exact gaze point of the eye by "corneal-reflection/pupil-center" method (Goldberg & Wichansky, 2003; as cited in Ball & Poole 2006: 211). For accurate measures of the point of regard, either the subject's head must be fixed, so that "the eye's position relative to [both] the head and point of regard coincide" (Duchowski 2007: 54; as cited in Tycová, 2015), or several features of the eye, such as corneal reflections, iris-sclera boundary and pupil shape, must be considered in order to dissociate the head and eye movements (Tycová, 2015).

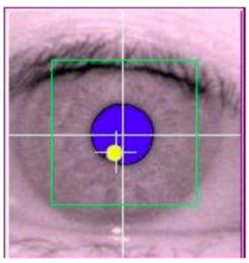

Figure 1 Pupil (blue disc) and corneal reflection (yellow dot and small cross) are identified by the eye tracking system. Photo credits to EyeLink 1000 manual.

Typical systems of this category usually comprise of a standard desktop computer, on which the stimulus is presented with an infrared camera mounted beneath (or next to) the screen, and image processing software to locate and identify eye movements. In operation, infrared light from a LED embedded on the camera is directed into the eye to create a strong reflection on the pupil. Upon entering the retina, a large proportion of the light is reflected back, while the corneal reflection is also generated, appearing as a small glint (Ball and Poole 2006: 211). The vector between pupil center and corneal reflection is then computed for the point of regard or gaze direction. If no head rest is employed, further head position calculations are required in computing the point of regard. For instance, EyeLink locates the subject with a target sticker on his/her forehead.

Since most eye tracking researches require information of the subject's point of gaze responding to a display of visual stimulus, system calibration is carried out to determine the correspondence between pupil position in the camera image and gaze position on the display. After calibration, validation shall be performed to identify the difference between target position. Gaze position is computed according to validation result to ensure the recording quality by correcting gaze-position inaccuracies which largely derive from errors in fixation data collected after the calibration.

Aside from the working mechanism of trackers, several technical properties, including sampling rate, accuracy and precision, are determining factors for certain research purposes, and should be taken into consideration in the discussion of methodology. The number of samples captured by an eye tracker per second is expressed by its sampling rate. Most modern eye trackers have a sampling rate ranging from 25 – 2000 Hz (Andersson et al., 2010). A 50Hz eye tracker registers a sample every 20 milliseconds, whereas the 1000Hz sampling rate gives a higher temporal resolution of 1 millisecond. The particular sample captured at a point is considered the representative of the whole interval of time, even though during that interval the eye does not stay in the same position. In this fashion, temporal sampling errors may occur. The temporal sampling error, as termed by Andersson, Holmqvist and Nyström (2010), refers to the time between the point of actual objective occurrence of an event and the detected occurrence of an event. With regard to this problem, different sampling rate should be adopted in accordance with distinct research purposes. For instance, researches involving gaze-contingent display changes require demanding sampling rate in order to detect saccade launches earlier and provide timely display changes.

Accuracy, another important technical issue in eye tracking, refers to “the average difference between what the eye-tracker records as the gaze position and what the gaze position actually is” (Bojko, 2013) and is measured in degrees of visual angle. Since the gaze is calculated by the features and reflections of the eye, a reference point, typically represented by pre-experimental calibration points, is required as the actual target by the system. In this regard, inaccuracies can be caused by, for example, glasses, contact lenses, heavy mascara and so on, which will be detrimental to position-based data analysis and interaction (Nyström et al., 2013), and thus need to be taken into account for decision of the size of an interest area. Consistent high accuracy of data throughout an experiment session in this project might be questionable, since hand or head movements during translation might impair the calculation of gaze position in calibration. Difficulties concerning experimental design will be discussed in details in the next section.

Precision, on the other hand, is defined as “a measure of how well the eye tracker is able to reliably reproduce a measurement,” which allows demanding standards to investigate the imperfections relating to oculomotor systems (Nyström et al., 2013: 2).

To sum up, the mechanism and technical properties of eye tracking systems may cause potential detriments to the validity of collected data, and thus should be put into consideration during the experimental design and data collection process. On the experimental design of this project, however, difficulties have been identified between high accuracy with head mount system and the validity of data due to unfamiliar environment to the participants. More details upon experimental design will be discussed in the next chapter.

2.3 Gaze Data as the Indicator of Attention

Data collected from eye tracking experiments has been argued to indicate various aspects of cognitive activities. In retrospect, a number of researchers have dedicated into forming a theoretical framework of the relationship between eye movements and cognitive effort. Towards a reading model, Just and Carpenter (1980) formulated "eye-mind hypothesis", assuming that "there is no appreciable lag between what is being fixated and what is being processed", and the immediacy assumption, stating that "a reader tries to interpret each content word of a text as it is encountered". The hypothesis speculates that when a particular word is fixated, no other word can be processed simultaneously until full comprehension is reached. Later works on eye tracking have frequently quoted this hypothesis, of which some arguments have been considered problematic.

Rayner (1998), in his comprehensive review of eye-tracking research, identified a strong connection between the location of fixation and cognitive effort, while his more recent works indicated that "[t]he eye and the mind are not tightly synchronized; the mind is sometimes a bit ahead of the eyes, but can also lag a little behind" and that "while fixating a particular word, there is a substantial amount of pre-processing of the next word" (Radach, Kennedy & Reynar, 2004), and obviously deviated from Just and Carpenter's hypothesis.

The dissociation of "spotlight of attention" from fixation suggests that it is possible to divert the attention to somewhere else while fixating a location (Posner, Snyder, & Davidson, 1980; as cited in Duchowski, 2002), such as the acquisition of information from parafoveal vision. It has posed a problem to eye-tracking researchers, since we cannot be fully confident that the fixated area is fully perceived. Some researches called for combined eye-tracking and brain-imaging equipment for a more comprehensive research in attentional tasks. However, it has not prevailed owing to prohibitive cost. In this regard, most eye-tracking researches still proclaimed that the eye trackers provide quantitative and objective evidence of asubject's visual and (overt) attentional processes (Duchowski, 2002).

2.4 Variables and Their Interpretation

Streams of raw data from aeye tracker are available after experiments, from which eye movement events are to be detected, either manually or automatically by software. Usually, event detection is operated by software automatically by applying a detection algorithm to the gaze data (raw data), so as to classify the data into various events, including fixations, saccades, blinks, smooth pursuit points and artefacts (noise). According to Holmqvist et al. (2011: 151), algorithms of this kind make use of three different streams of data: gaze coordinates, gaze velocity and gaze acceleration (as cited in Tycová, 2015). These events may reflect emotional states and cognitive processes, thereby being said to capture both conscious and unconscious dimensions of a subject's behaviors. The successive paragraphs address several common events in eye tracking research via discussion of their features, general principles of event detection as listed by Holmqvist et al. (2011) and theoretical interpretations. The measures implemented in this project are also included to facilitate the presentation of experimental results later in a data analysis section.

2.4.1 Fixations

Poole and Ball (2005) describe fixations as "moments when the eyes are relatively stationary". Instead of representing a single actual fixation of the eyes, fixations in the sense of eye tracking are predominantly detected by a maximum allowed dispersion or by velocity. The adjacent samples of raw data would be computed as fixations where, in the former method, they fall in limited region for a minimum duration, or in the latter method, the gaze velocity does not exceed a predefined threshold (Tycová, 2015).

The interpretation of fixations varies greatly according to various contexts. In encoding tasks, such as webpage browsing or source text comprehension during translating, higher fixation frequency on a particular area may indicate a greater interest in the target, or greater difficulty to encode. Additionally, the duration of a fixation is also said to be associated with the processing time applied to the target, and therefore, implies the difficulty in extracting information (Just & Carpenter, 1976).

Dwell time, or the sum of all fixation durations within an interest area, is best used to compare attention distributed between targets, or "as a measure of anticipation in situation awareness if longer gazes fall on an area of interest before a possible event occurring" (Poole & Ball, 2005).

2.4.2 Saccades

Saccades, detected by velocity and acceleration of eye movements (SR Research, 2005), refer to short and quick movements between two fixations. Sensitivity to visual input is reduced during saccadic movements (Rayner, 1998). However, regressive saccades, or regressions, are usually interpreted as the processing difficulty during encoding. While most regressions are small, larger regressions may indicate in higher-level processing of the text (Rayner & Pollatsek, 1989; as cited in Poole & Ball, 2005), for instance, the syntactic error in machine translation output. Larger saccade amplitude can also represent more meaningful cues, which draw attention from a distance (Poole & Ball, 2005). Further, large directional shifts of saccades, especially those exceeding 90 degrees, may manifest that a user's goal has changed, or the interface does not abide to the user's expectations in user experience design (Cowen et al., 2002).

2.4.3 Blinks and Pupil Size

Blinks are detected through partial occlusion of the pupil, causing artificial changes in pupil position, and the duration of missing pupils. These two measures usually function as an indicator of cognitive workload. A lower blink rate is assumed to imply a higher workload, and a higher blink rate may indicate fatigue (Bruneau, Sasse, & McCarthy, 2002; Brookings, Wilson, & Swain, 1996; as cited in Poole & Ball, 2005). Larger pupil sizes may also indicate emotional arousal or more cognitive resources required for processing. Other factors, such as lighting, may also cause dilation or constriction of pupils.

2.4.4 Area of Interest (AOI)

AOIs can be defined by researchers as specific areas within which they want to analyze and compare eye movements. For example, in usability tests, well-defined AOIs allow researchers to single out certain content or components of the interface (Bergstrom & Schall, 2014), and analyze altogether with the aforesaid measures to identify the user experience of a certain visual stimulus.

2.4.5 Heat map

Based on fixation counts, the visualization of eye movements serves as an aid in understanding where the participants fixate the most. This kind of visualization also enhances the communication with the clients than mere figures.

3. Experiment Design

In this section, various aspects of experimental design in this project are presented and explained. Difficulties encountered in design are also identified.

3.1 Research Questions

The project aims to investigate the distribution pattern of cognitive workload across translation subtasks via eye-tracking experiments. While this project is not intended to address all aspects of the translation process and is also not restricted to only one single question for study, a few research questions are identified at first place as follows:

- What is the general distribution of cognitive effort during translation process that can be observed in terms of eye movements?
- How are cognitive effort distributed among subtasks of translation? Are there any identifiable eye movement patterns for each subtask or domain?
- What are the eye movement patterns across subtasks/domains?
- Can characteristic fixation sequences be observed (Zwierzchoń-Grabowska, 2011)?
- To what extent will the text complexity affect different measures of eye movements during translation process?
- Is translation conducted in a sequential fashion as suggested by Gile (2011) or any overlapping process can be identified (Hvelplund, 2011)? To what extent does such parallel processing take place?

3.2 Experiment Settings

The translation task was performed with a computer in a windowless room with only artificial lighting. Participants were asked to conduct the translation with their familiar text input methods. The presentation of the screen was divided into three parts, namely ST, TT and the dictionary, with fixed boundary for the ease of data analysis. The layout made reference to the working environment of commercial computer-assisted tools, such as Atril Déjà Vu X2.

The layout was based on HTML webpage for the ease of identifying AOIs, and a primitive input area was given as the editing environment. The font size of 16px and the serif font Georgia in color #333 (dark grey) were adopted for the display with 150% line spacing. The display of the source text made reference to online news pages, including New York Times and The Guardian, which opt for proven presentations for efficient reading in existing research (Beymer et el, 2008). Additionally, the layout also referenced the typical font size of 16-20 as suggested by O'Brien (2009) for research, which allows in-text analysis.

A head-mounted eye tracker was adopted for the experiment, and inevitably, the participants might be influenced by the physical constraints of the setting which did not completely resemble their usual working environment. In this regard, reasonable text length and enough break for the participants were considered in the experimental design.

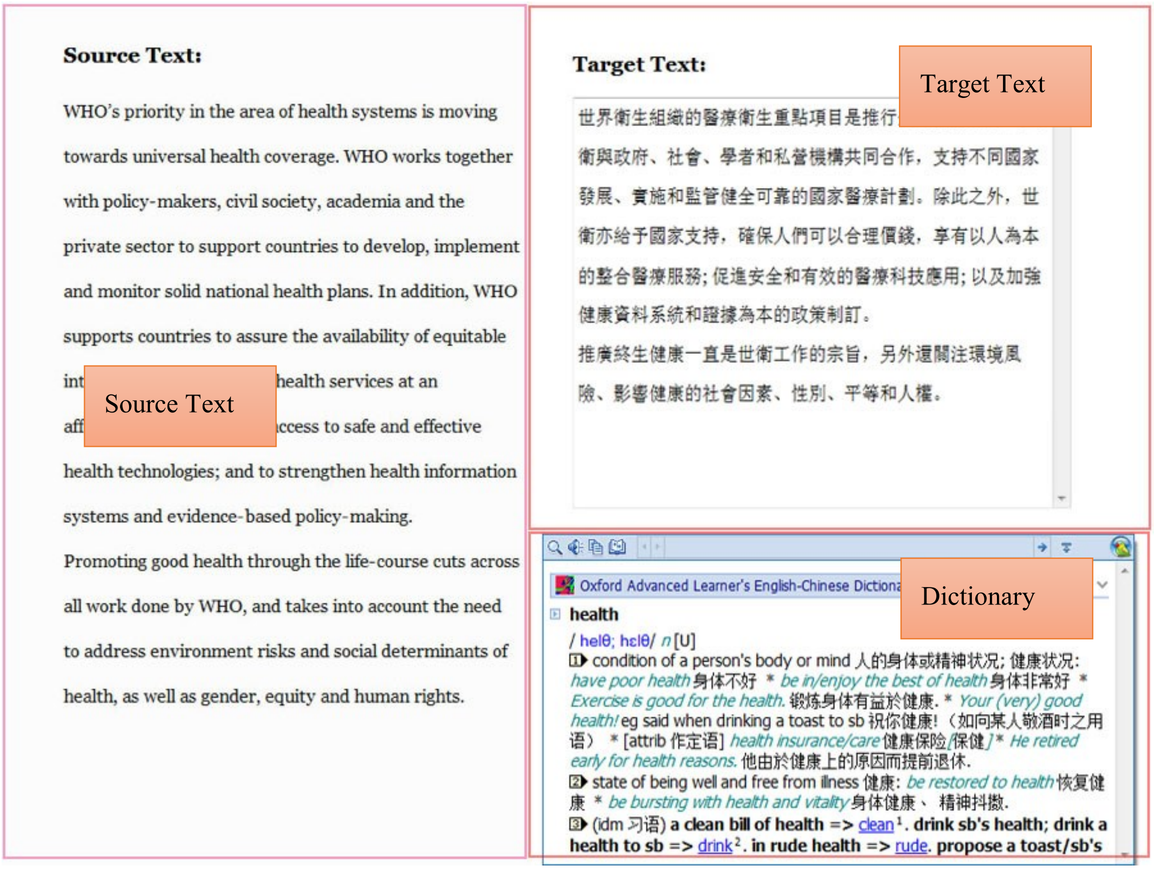

Figure 2 Display layout. ST, TT and the dictionary were presented simultaneously. Red borders and interest area labels were added here for clarification.

3.3 Scope of Participants

In this project, the scope of participants targeted to translation major students who had completed more than a year of their study with Chinese, either Mandarin or Cantonese, as the native language. As required by the experiment nature, a participant should have normal or corrected-to-normal vision.

14 participants were recruited while data from 11 participants were adopted for data analysis. All valid data sets were collected from final-year undergraduate students of the translation major in City University of Hong Kong.

3.4 Experimental Texts

Experimental texts (please refers to the Appendix for whole texts) for the experiments in this project were of general purpose, of a similar length in terms of word count but of variant complexity. The then decided texts were the abridged version from the introduction page of international organizations, which regard general public as the target audience. No further modifications were made to the sentences. Each text consists of approximately 110 words. A text longer than this would require the participants to scroll the ST window upon reading, and inevitably increase the possibility of the participants being uncomfortable with physical constraints.

Indicators of text complexity proposed by Jensen (2011) were introduced during the selection of experimental texts, by which the relative differences in complexity were measured by three quantitative criteria. Such indicators, including readability indices, word frequency and non-literalness, were suggested to objectively assess the relative amount of comprehension and production efforts needed during a translation process, and were incorporated in this project as a variable of relative attentional efforts.

Readability index, mostly based on quantifiable properties such as syllables, words and sentences, is arguably claimed to indicate the ease of comprehending a text. All readability indices adopted indicated that Text 1 has higher level of complexity. For example, from the statistics, the U.S. grade level indices revealed that 18.9 years of schooling were required to comprehend Text 1, while 9.4 years were needed for Text 2.

Regarding the second parameter, word frequency was calculated with Corpus of Contemporary American English (COCA), while lemmas were counted for different word forms in a family. Similar proportions of low- and high-frequency words were found in two texts, with Text 1 having slightly higher percentage of low frequency words.

Non-literalness does not take significant presence in the experimental texts due to the text nature. Therefore, it was not regarded as a major affecting variant in this project. Although the aforesaid complexity indicators proposed a quantitative method to assess text complexity, the level of difficulty of a text may prove problematic to gauge, that is, a complex text is not necessarily difficult to translate, as it depends very much on the routines, competence and specializations of the translator (Jensen, 2010). However, since a complex text generally indicates higher difficulty of comprehension and translation, these relatively crude measures can still serve as a rather objective method to assess the relative difficulty of texts. Therefore, according to Figure 3 and 4, Text 1 is considered to be more complex and hence assumed to require more efforts for processing than Text 2.

	Flesch Kincaid Reading Ease	Flesch Kincaid Grade Level	Automated Readability Index (US Grade)
Text 1	23.9	17.1	18.9
Text 2	63.2	8.2	9.4

Figure 3 Readability indices of experimental texts.

	Word Count	Fq: 1-500	501-3000
Text 1	110	57%	20%
Text 2	105	64%	15%

Figure 4 Word frequency in distinctive frequency ranges. 1-500 accounts for common words such as like, the, etc., while >3000 stands for low frequency words.

3.5 Task

The participants were asked to translate one warm-up text and two experimental texts respectively from English to Chinese, while the display sequence of the two texts were counterbalanced among participants. One offline dictionary was made available simultaneously on the monitor as one of the three interest areas. No time constraints were implemented for the tasks, and the participants were told to translate up to their usual quality criteria without any mandatory procedure being required. .

3.6 Difficulties

While the experimental design strived to consider all affecting factors, influence from physical settings and variations of participant profile were uncontrollable factors that might make the validity of data questionable.

Firstly, although problems concerning eyes and glasses could be anticipated and taken into account to ensure successful recording, certain optical aids might still jeopardize the validity or quality of the data collection. Therefore, some data so collected would be of poor quality or even need to be discarded. For example, 1 of the 14 participants in this project was founded to wear blue light-blocking glasses that crippled the quality due to the working mechanism of the eye tracker. Other conditions, such as bifocal lenses, narrow eyeglass frame or heavy mascara, would possibly affect the recording quality as well (O'Brien, 2009). They were avoided by instructions or not encountered in this project.

Further, human bias takes a significant role and often generates debatable results in translation studies. For example, a limited size of participant group may lead to debatable conclusions. In a discussion on eye-tracking methodology, O'Brien (2009) argued that the average number of 12 participants in eye-tracking studies of translation made the validity of generalizations questionable. A larger participant group was recommended in researches aiming to render a generalized translation process. On the other hand, the levels of translation competence among participants are also a factor very difficult to monitor, and additionally, their backgrounds or specializations may also influence their translation behaviors and performance.

Few aspects of human bias proposed by Campbell and Stanley (as cited in White, 2008) on translation evaluation could be borrowed in discussion here. On the one hand, history effect indicates that events outside the world of one's judgments can intervene, which may lead to inconsistent or unreliable judgments, including linguistic judgments, among participants and hence inaccurate conclusions about the results. On the other hand, maturation effect suggests the very ordinary but principal factor that the participants might get tired, bored, or fed up with the process of the task can influence their behaviors and performance. Last but not least, the personal quality standards, which each participant was advised to achieve in experiment, varied inevitably and were hard to monitor, and hence would bring individual differences in to the results.

In conclusion, experimental design involves considerations of various factors and aspects, including variables for analysis, the physical environment, humans and so on, in order to obtain valid and useful results. Difficulties encountered in this project, as discussed above, may reveal inevitable limitations of this project and suggest directions of improvement for possible future research.

4. Data Analysis

Eye-tracking data was collected from experiment sessions to investigate the distribution of cognitive effort and workload of translators during translation. In total, 14 participants were recruited, and 3 sets of data were discarded due to poor quality

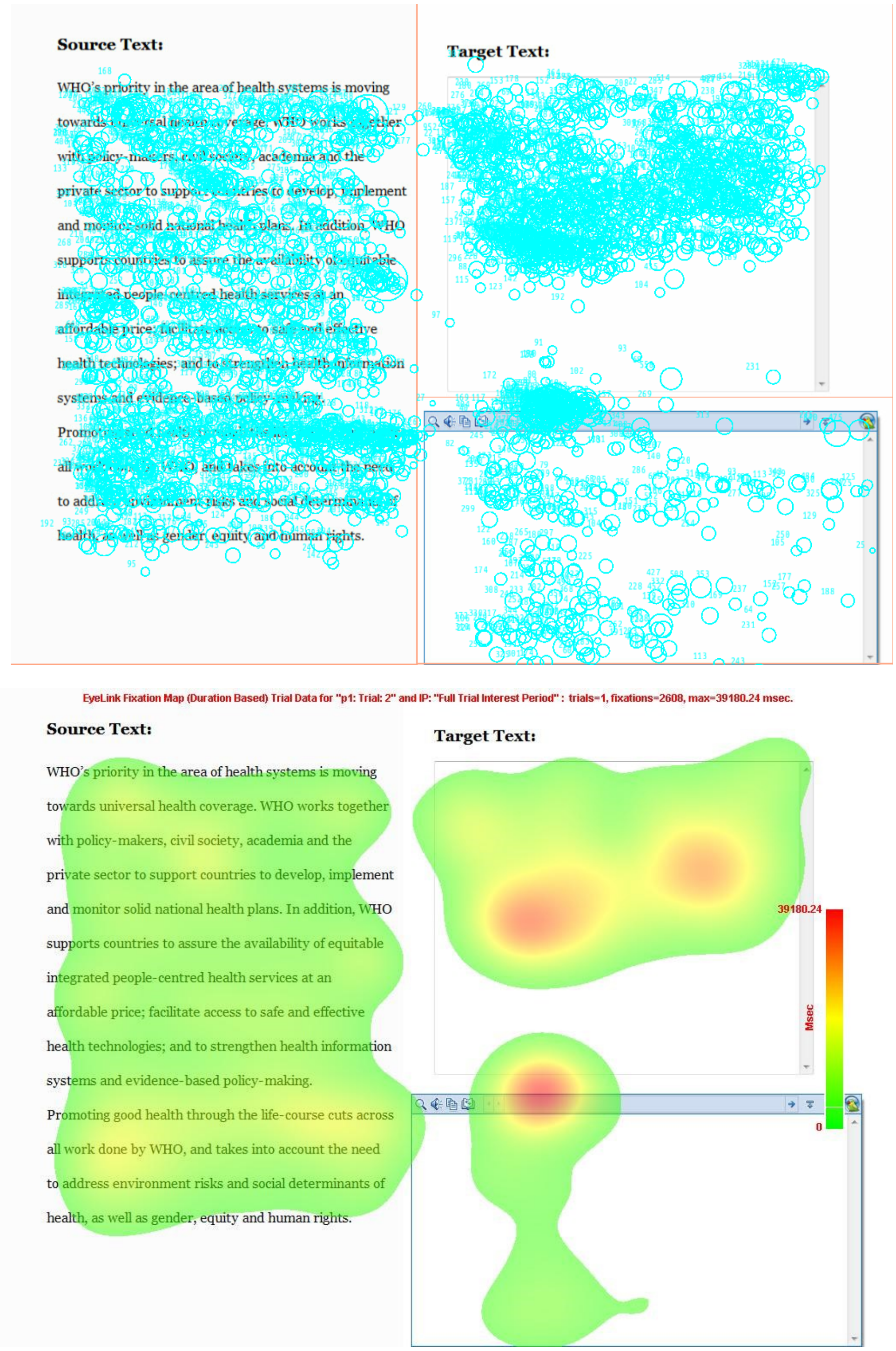

Figure 5 Sample of visualized eye-tracking data of a single participant. The above showcases fixation points, of which the diameter indicates the duration and the below displays the heat map of the session.

caused by optical aids and/or the input method, Qcode, which induced on-screen distractions to eye movements.

Striving to provide a comprehensive description on different aspects of translation process, this project analyzed various measurements of eye-tracking data. Mainly three aspects of translation process were studied, including (1) the general distribution patterns of cognitive effort, (2) levels of cognitive workload among different subtasks during translation and (3) working style of translation. Two independent variables, Text Type (refers to as Text 1 and Text 2, indicating different levels of complexity) and categories of working tasks, which will be discussed in details in the following part, were incorporated to analyze their relationship and effects with the aid of various measures of eye-tracking data.

4.1 Categorizing Subtasks

Two ways of categorizing translation subtasks were adopted in this project, and termed as AOIs and Subtask Type hereinafter. AOIs were identified automatically with fixations in predefined boundaries on screen, which were basically the three areas, ST, TT and dictionary as illustrated in the previous section, and were considered to indicate the current working interest of a participant. No further analysis was conducted to separate the subtasks within AOIs.

On the other hand, eye-tracking data was analyzed together with key-logging data, and accordingly, a few subtasks were identified. Four subtasks were first observed and categorized as ST comprehension, TT production, dictionary lookup and parallel attention (PA, termed by Jensen (2011), implying the simultaneous processing of comprehension and production) by aligning the AOIs and keys in adjacent rows. This type of categorization basically reinterpreted the fixations in ST area. ST comprehension refers to fixations on ST area while no typing events are detected; on the contrary, if a typing event is detected while the subject fixates on ST area, it will be coded as PA. TT production and dictionary lookup stands for fixations on these respective areas, while key events with no interest area information, interpreted as fixations on the keyboard, will be coded into TT. If no match for both

AOIs and key-event for a fixation is found, then the data will be coded as “no data” and will not be used in data analysis procedure.

However, some limitations of such categorization need to be identified in advance. Argued by Jensen (2011), as questionable examples of PA may be registered due to a delay of at least 180ms for typing event, it may pose a risk of registering short typing activities that are observed to occur simultaneously with ST comprehension without consideration of the delay. Additionally, since the registration of PA relies on key events, it is not able to tell whether there are cognitive activities of production in mind without typing that take place during ST comprehension. Despite these drawbacks, categorizing fixations and key events by subtasks still demonstrates discrete merits, and provides a closer look into details of the translation process than AOIs categorization alone.

4.2 Overall Distribution

The dwell time of a distinctive category was analyzed in accordance with Text Type and different subtask categories to compare the levels of cognitive attention distributed to different targets and render a generalized picture of translation process.

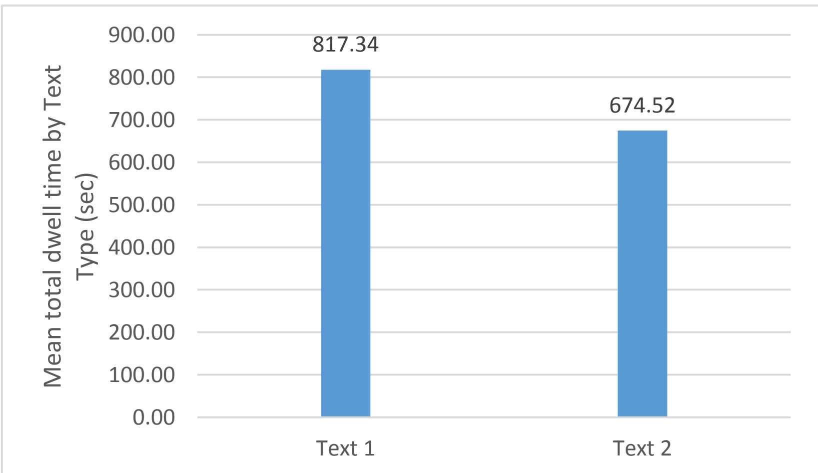

Figure 6 Mean total dwell time by Text Type (seconds).

4.2.1 Dwell time × AOIs

Dwell time, as mentioned in Section 2, presents the temporal length that a translator stays in a certain area of task, and is argued to stand for the cognitive attention devoted to that very task. Figure 6 illustrates the total dwell time on respective texts in absolute mean values. The mean dwell time was 817.34 seconds for Text 1 and 674.52 seconds for Text 2, in conformance with their text complexity. In both texts, as indicated in Figure 7, more attention was directed to the TT during

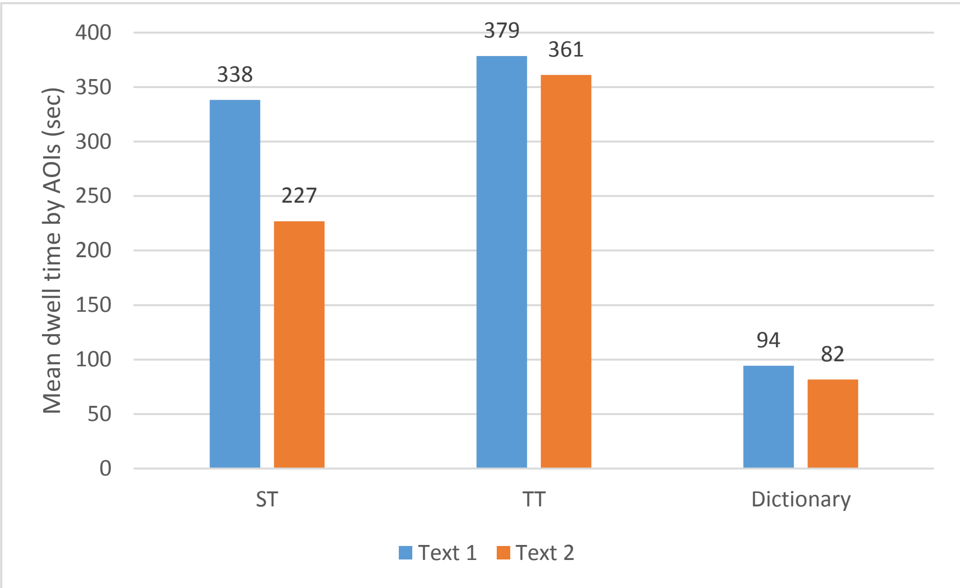

Figure 7 Mean dwell time by Text Type in different AOIs (seconds).

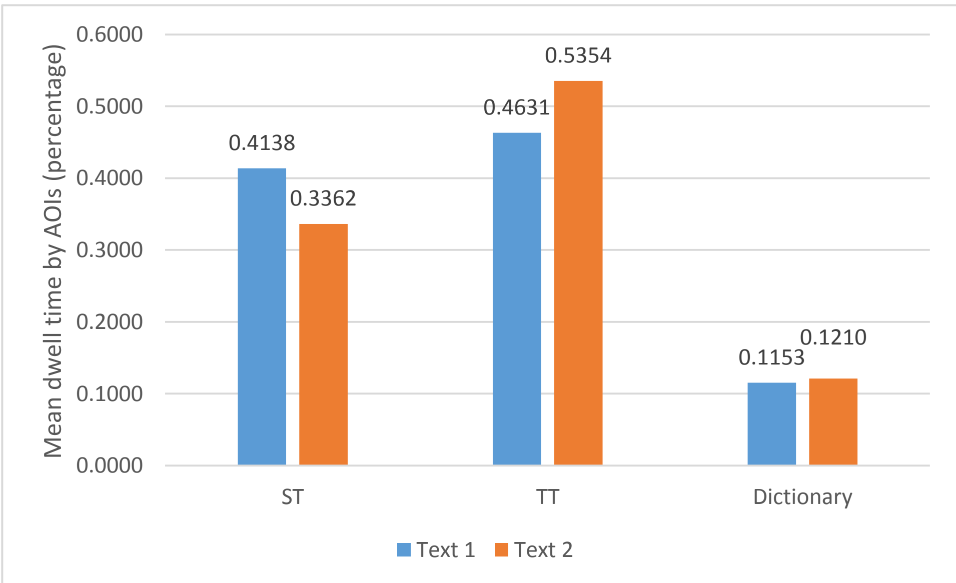

Figure 8 Percentage distributions of dwell time by Text Type in different AOIs.

translation for both texts. The means were found to be 378.50 seconds and 361.14 seconds respectively. ST areas received the means of 338.22 and 226.78 seconds, and additionally, as Text 1 scored higher in text complexity, proportionately more attention was found in comprehension of ST for Text 1, as shown in Figure 8. It is revealed by statistics that Text 2 demanded less cognitive effort in text comprehension than Text 1 but a relatively greater proportion of efforts were found in production than in comprehension. However, a similar percentage of attention was dedicated to the dictionary in both texts, corresponding to 11.53 and 12.10 percent on the two texts respectively.

Two-way analysis of variance (ANOVA) was conducted to analyze to what extent dwell time was influenced by Text Type and AOIs. The result showed main effects of AOIs on dwell time ($F=16.749$, $p<0.001$), whilst Text Type ($F=1.340$, $p=0.252$) and their interaction ($F=0.625$, $p=0.539$) manifested no significance, suggesting that the distribution patterns among different AOIs during translation remained relatively constant between two texts, and consequently, AOIs served as a major factor of the distribution. Concerning the experimental design of this project, it needs to be noted that this outcome is only applicable to a translation task without time constraints so far.

A Turkey HSD post-hoc test on AOIs showed that, except for ST-TT pair ($p=0.196$), all other pairwise differences between AOIs were significant (ST-Dictionary and TT-dictionary: p’s<0.001 for both). Further t-test analysis within each text indicated that the ST-TT pair was not significant as well (Text 1: $F=2.852$, $p=0.645$; Text 2: $F=4.394$, $p=0.109$), lending no support to the assumption in the descriptive analysis that lower text complexity resulted in relatively (i.e., proportionally) longer attention distributed to TT area, since no main effect was found from the Text Type. Whether attention distribution between ST and TT has anything to do with their text complexity in a significant way is yet to be further examined.

4.2.2 Dwell time × Subtask Type

Dwell time by Subtask Type was analyzed to investigate the attention distribution among different subtasks. As mentioned in the previous paragraphs, the fixations with typing events, which landed in ST area, were lining out and marked as Parallel Attention (PA), whereas TT and Dictionary lookup remained constant. Figure 9 presents the dwell time for subtasks on different texts. ST comprehension received 297.99 seconds and 194.78 seconds respectively, and PA recorded 34.25 seconds and 26.32 seconds. Considering the proportion of dwell time allocated to each subtask, as

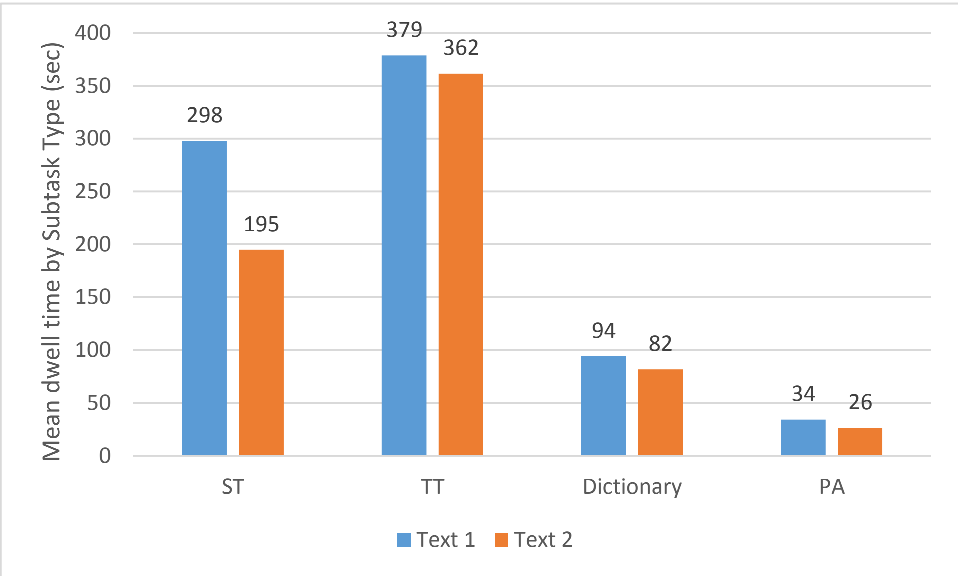

Figure 9 Mean dwell time by Text Type in different subtasks (seconds).

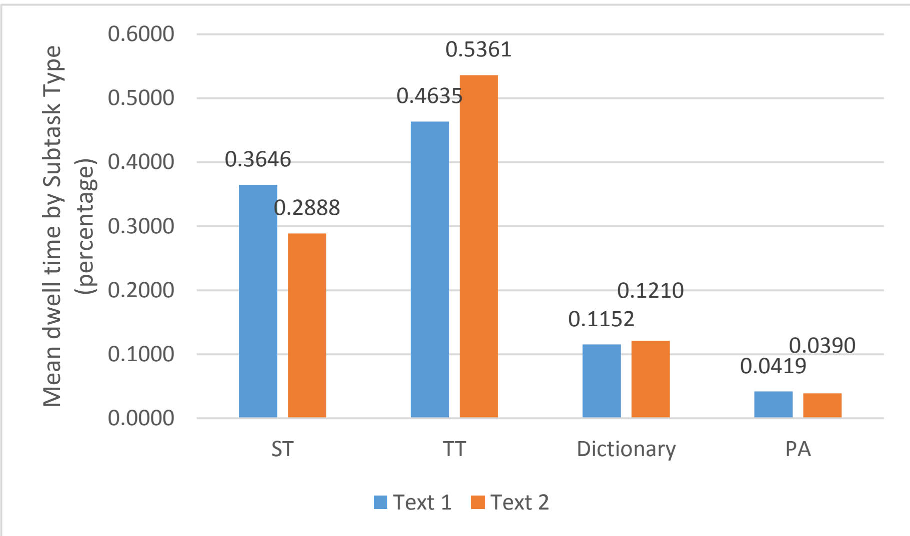

Figure 10 Percentage distributions of dwell time by Text Type in different subtasks.

illustrated in Figure 10, TT production still accounted for proportionately significant attention. While Text 1 received higher percentage of dwell time in ST comprehension than Text 2, similar percentages of PA, 4.19% and 3.90% respectively, were observed one the two texts, with Text 1 slightly higher.

The result of two-way ANOVA signified main effect of Subtask Type ($F=108.831$, $p<0.001$), while no significance was found in Text Type ($F=0.553$, $p=0.459$) and the interaction ($F=0.900$, $p=0.445$). In addition, post-hoc analyses displayed main effects of Subtask Type in all pairs (p's<0.001 for all) except for Dictionary-PA ($p=0.176$). Unlike AOIs categorization, ST-TT pair showed significance in Text 2 ($F=4.633$, $p=0.045$) but not in Text 1 ($F=2.681$, $p=0.119$). The result indicated that, after re-categorizing the subtasks in ST area, a statistically significant increase of attention was directed to TT production comparing with comprehension in Text 2. Further analysis was conducted to identify whether Text Type has main effect to dwell time in PA subtask. The results turned out to be no main effect of Text Type ($F=3.665$, $p=0.454$), suggesting that there was no significant difference caused by text complexity in dwell time of PA.

Merely relying on dwell time, nonetheless, might not be sufficient for analyzing the cognitive effort required in different aspects of translation process. Hence, the fixation duration and pupil size, representing the time of processing and the amount of cognitive resources required for processing information, were analyzed for further discussions.

4.3 Cognitive Attention by Fixation Duration

4.3.1 Fixation duration × AOIs

Two-way ANOVA of fixation duration was conducted to analyze the cognitive workload in distinct AOIs. It turns out that AOIs had main effect ($F=402.184$, $p<0.001$), and additionally, according to results of the post-hoc analysis, all pairwise AOIs showed strong significance (p's<0.001 for all pairs), indicating there were statistically significant differences in the duration of fixations between three interest areas. On the other hand, Text Type ($F=0.505$, $p=0.477$) and the interaction between Text Type and AOIs ($F=2.945$, $p=0.053$) manifested no significance. Figure 11 shows that fixations in TT accounted for the longest mean durations (275.04 and 272.91 by Text Type respectively), while dictionary area presented longer mean durations than ST area.

T-test with Text Type as an independent factor was also conducted. Significant difference was found only in ST ($F=12.348$, $p<0.001$), not in TT ($F=0.752$, $p=0.386$) and dictionary ($F=1.490$, $p=0.222$). It could be inferred that text complexity had main effect on the fixation duration when the fixations landed in ST, and higher complexity led to longer durations.

4.3.2 Fixation duration × Subtask Type

Two-way ANOVA with Text Type and Subtask Type as fixed factors was conducted as well. As a result, Subtask Type ($F=290.429$, $p<0.001$) and the interaction of Subtask and Text Type ($F=2.602$, $p=0.05$) demonstrated significant effects, while Text Type ($F=3.682$, $p=0.055$) alone failed to reach any significance.

Figure 12 illustrates the means of fixation durations with Subtask Type. The main effect of Text Type within each subtask was examined. Similar to AOIs, no significant differences were found in TT ($F=0.556$, $p=0.456$) and dictionary lookup ($F=1.490$, $p=0.222$). On the other hand, the increase of durations by text complexity was significant in ST ($F=10.614$, $p=0.001$) and PA ($F=4.213$, $p=0.040$). This

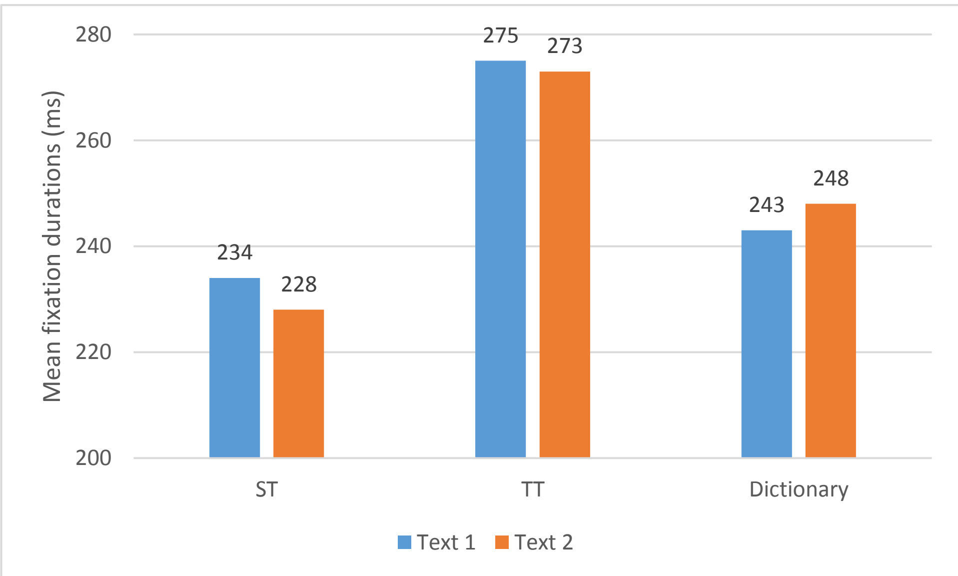

Figure 11 Mean fixation durations by Text Type in different AOIs (seconds).

reconfirmed the result from the analysis of AOIs that higher text complexity led to greater durations, and the result from the analysis of Subtask Type that higher text complexity required more cognitive effort in ST comprehension and parallel processing of comprehension and production.

According to the figures so far, it could be assumed that subtasks containing production, including TT production (274.84 and 273.01 ms) and PA (278.23 and 264.04 ms), required greater cognitive attention than ST comprehension and dictionary lookup. Following on this assumption, post-hoc analyses were carried out to examine pairwise significance. Main effect of Subtask Type was discovered in all pairs of ST-TT ($F=206.563$, $p<0.001$), ST-PA ($F=846.247$, $p<0.001$), Dictionary-TT ($F=119.333$, $p<0.001$) and Dictionary-PA ($F=40.764$, $P<0.001$), lending further empirical support to the assumption. The difference between TT and PA was, nonetheless, insignificant by all factors, including Text Type ($F=3.303$, $P=0.069$), Subtask Type ($F=0.401$, $p=0.526$) and their interaction ($F=1.966$, $p=0.161$). Further analysis needs to be deployed in order to identify the relationship of cognitive effort in TT and PA.

The result also signified that even the fixations were found in the same interest areas, there were distinctive subtasks requiring different levels of cognitive attention during translation. Additionally, the presence and significance of PA that

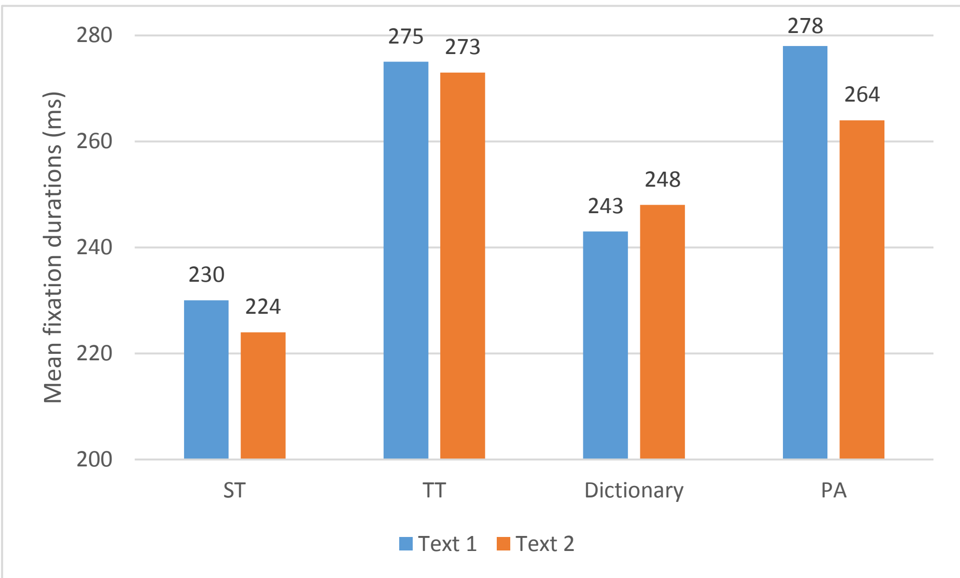

Figure 12 Mean fixation durations by Text Type in different subtasks (seconds).

characterized the simultaneous activation of comprehension and production efforts bring up a counterargument to the vertical view of translation that translation is not necessarily an exclusively serial processing task. Further analysis of pupil size will be conducted next to test this argument and investigate the unanswered question about the level of cognitive effort differences between TT and PA.

4.4 Cognitive Attention by Pupil Size

4.4.1 Pupil Size × AOIs

Pupil size was measured as an indicator of cognitive workload as well. Similar analysis was conducted on it as for fixation durations by setting Text Type and AOIs as fixed factor in comparison with Text Type and Subtask Type. The data was presented in arbitrary units subjected to the influence of 10% from pupil position. Typical pupil area falls between 800 to 2000 units (SR Research, 2005).

By categorizing fixations with AOIs, the result presented main effects of Text Type (F=620.038, p<0.001) and AOIs (F=144.867, p<0.001), while the interaction was not found significant (F=2.111, p=0.121). These results indicated that the manipulation of text complexity affected pupil size. As shown in Figure 13, pupil size of Text 1 was clearly greater than Text 2 in all AOIs, implying a higher level of cognitive workload was demanded for the more complex text, namely Text 1.

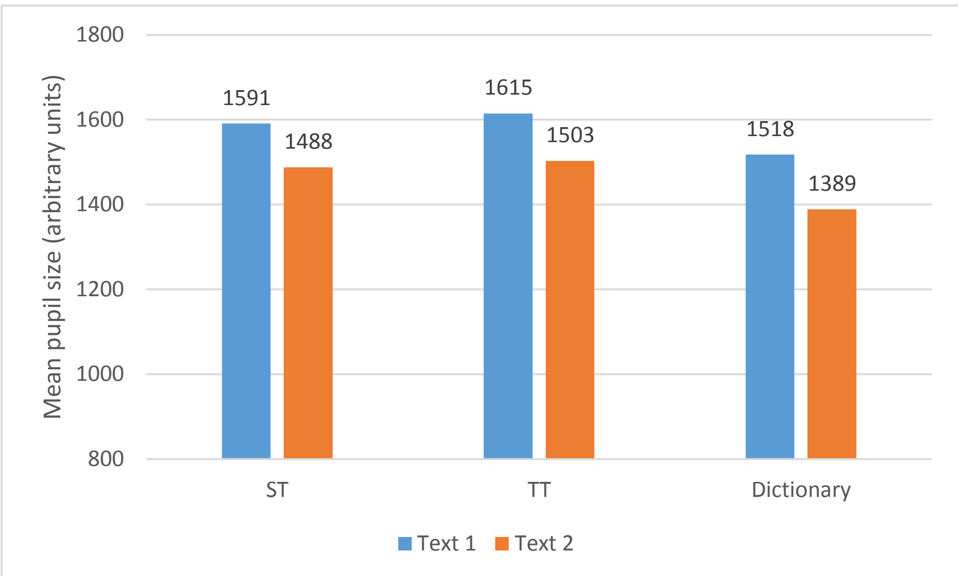

Figure 13 Mean pupil size by Text Type in different AOIs (arbitrary units).

Post-hoc analysis was conducted to examine the pairwise significance between AOIs. The result manifested significant differences among all pairs (ST-TT: p=0.023; other pairs: p's<0.001), and consequently, suggested that TT area received significantly larger cognitive workload, whilst dictionary area the least.

4.4.2 Pupil Size × Subtask Type

When the data was categorized into subtasks, the pupil size displayed strong significance in all factors, including Text Type (F=221.758, p<0.001), Subtask Type (F=102.118, p<0.001) and the interaction of the two (F=10.147, p<0.001), indicating that all these factors would pose effects to the dilation or contraction of pupils. In addition, post-hoc pairwise analysis demonstrated significance in all pairs (Dictionary-PA: 0.002; other pairs: p's<0.001) except for ST-TT (p=0.614). Comparing with the contradictory finding in AOIs which discovered significance in ST-TT pair, the result indicated that, after excluding PA processing, ST comprehension (without typing events) and TT production actually had no significant difference in levels of cognitive effort. While the fixation duration analysis implied that parallel processing required more cognitive effort due to its nature of simultaneous processing, the result in this analysis, however, showed that PA demanded less cognitive effort than comprehension and production. Further discussions on this account will be given in the following part.

On the other hand, t-test with text complexity as an independent factor showed main effect on ST, TT and dictionary lookup (p's<0.001 for all three), except for PA

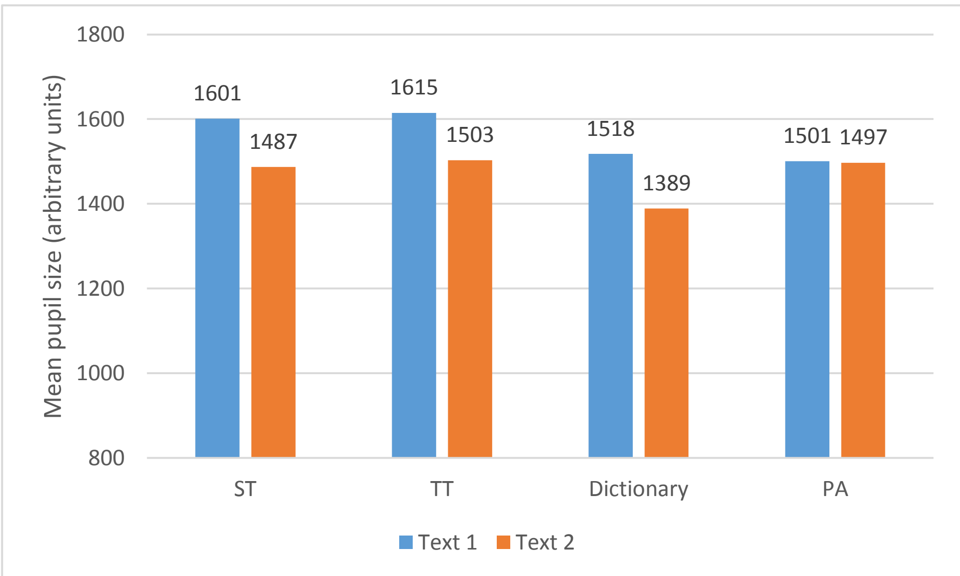

Figure 14 Mean pupil size by Text Type in different AOIs (arbitrary units).

(F=1.00, p=0.751). The result indicated that the cognitive effort required in PA would not be influenced by text complexity, which again, contradicted with the finding in the analysis of fixation duration.

The three subtasks, ST, TT and PA, were examined with extra attention to investigate the levels of cognitive effort of simultaneous processing comparing with independent subtasks, ST comprehension and TT production, with respect to text complexity. Pairwise analysis, ST-PA and TT-PA, were deployed and resulted in inconsistent relationship of pairs within respective texts. Firstly, in TT-PA pair, main effect was found from Subtask Type in Text 1 (F=47.778, p<0.001) but not in Text 2 (F=0.188, p=0.665). On the other hand, ST-PA pair, which was in assumption to display a significant increase in PA as suggested in fixation duration analysis, in fact displayed a significantly smaller pupil size for PA in Text 1 (F=46.839, p<0.001) while no significant difference in Text 2 (F=0.682, p=0.409) was discovered. Possible reasons of this phenomenon will be discussed as follows with support from the analysis of subtasks involved in dictionary lookup, and additionally, further explanations on the discrepancy of findings in two measures will be .

4.5 Discussion on Cognitive Workload

4.5.1 Key Events and Cognitive Workload Indicators

By integrating both the fixation duration and pupil size, cognitive workload of various subtasks in translation could be examined from a different perspective. The results, however, manifested inconsistency between the analysis of fixation duration and pupil size. In both categorizing methods, dictionary required longer fixation duration, albeit smaller mean pupil size was detected. Since the aforesaid analysis did not separate the search subtask, which consists of typing events, within dictionary, further analysis was conducted in this paragraph to verify the assumption that subtasks containing key events might lengthen the fixation durations.

A two-way ANOVA was deployed, and the presence of key events and Text Type were adopted as fixed factors. The result, as illustrated in Figure 15, confirmed the assumption that the presence of key event had main effect on fixation durations (F=238.83, p<0.001), while Text Type (F=1.163, p=0.281) and their interaction (F=0.457, p=0.499) did not contribute to the differences significantly. It indicated that the presence of key events would lead to longer fixation durations irrespective of text complexity.

Additionally, pupil size was also analyzed according to the presence of key event in dictionary task. The result demonstrated significant effects from all factors, including Text Type ($F=32.830$, $p<0.001$), the presence of key events ($F=27.180$, $p<0.001$) and the interaction of the two ($F=4.061$, $p=0.044$). As illustrated in Figure 16, as expected, the pupil size reflected and fluctuated with text complexity. And additionally, the decrease in pupil size for cases with key events indicated the possibility that less cognitive effort was required even if the fixation duration was longer.

In this regard, in the discussion of cognitive effort required during translation, pupil size may be a rather reliable indicator when key events are involved, otherwise,

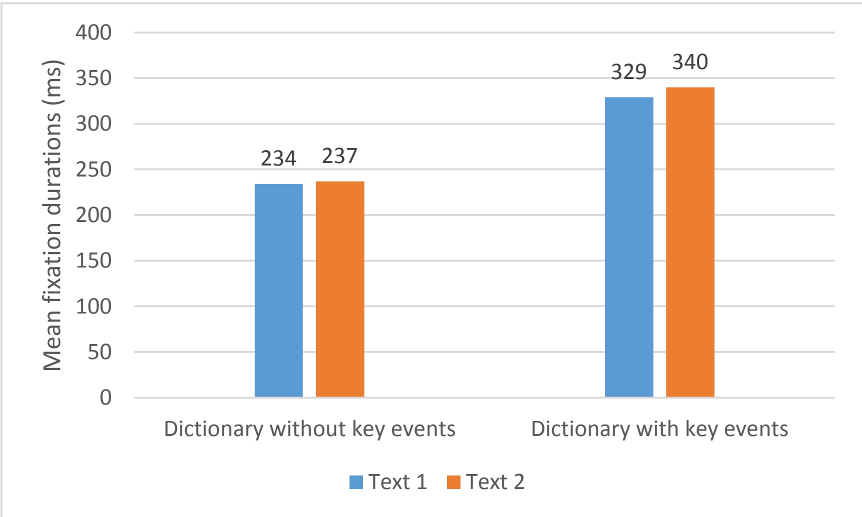

Figure 15 Mean durations of fixations in dictionary area with vs. without key events (seconds).

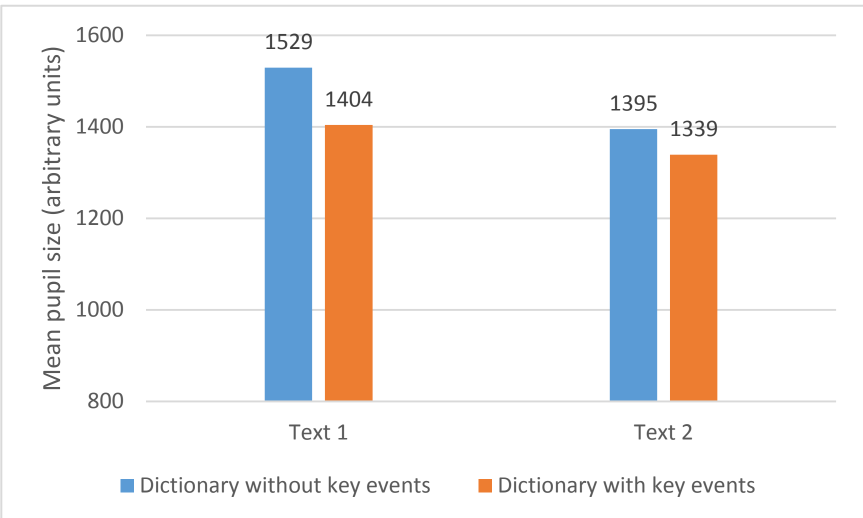

Figure 16 Mean pupil size in dictionary area with vs. without key events (arbitrary units).

the identification of subtasks should be done in a comprehensive and discreet manner in order to avoid misinterpretation of the data.

4.5.2 Discussions on PA Processing

The observation discussed in the previous paragraph could help explain the contradicting results from fixation duration and pupil size in PA processing. As discussed in 4.3.2, PA displayed significantly longer durations than ST in both texts, whereas statistically smaller (Text1) and similar (Text 2) mean values of pupil size were observed. Longer duration, as explained above, could result from typing events rather than only a greater cognitive workload.

On the other hand, PA also manifested inferior (Text 1) or similar (Text 2) values in both fixation duration and pupil size comparing with TT. Both subtasks involved typing events, similar or less cognitive effort was drawn to PA than TT production. It is straightforward to assume that parallel processing, which simultaneously tackle more than one task, required more cognitive workload than mere ST comprehension or TT production. A sound explanations is yet to be found for this phenomenon.

One explanation, as suggested by Hvelplund (2011), is that the measurement of PA does not reflect parallel processing but perhaps either ST or TT, that is, the categorization is problematic. However, this argument cannot explain the statistical significance of our data in Subtask Type and also pupil size for ST, TT and PA, as previously illustrated in Figure 14. Another explanation suggested by Hvelplund is automaticity of processing, proposing that only one task can be at the center of attention at any given time. It suggests that either ST or TT is processed consciously, and consequently, demands less cognitive resources than parallel processing. For instance, when a translator is comprehending and reconstructing ST messages, typing is automatically engaged. Similarly, whilst encoding translation output, a translator may also be engaged in ST reading, which is not cognitively demanding in the sense that it involves only orthographic analysis within sensory memory (Hvelplund, 2011).

Another more probable explanation has to do with the interpretation of the measure in use. It is reported in a variety of studies that the subject's pupillary dilation responses to "emotionally toned or interesting visual stimuli", according to Hess and Plott (1960). Pupil size in PA processing, which is an attention-splitting task, may only reflect the cognitive effort drawn for visual processing. However, this explanation has not obtained any empirical evidence as support from this project. Further studies, possibly with different methodologies, seem to be required to investigate this issue.

4.6 Saccades and Working Style

While most of the discussions above has been dedicated to fixations, how the translators switched between AOIs may also help to examine and explain their working style.

4.6.1 Shift Probabilities

Cross interest area saccades were computed into three categories according to their directions, including saccades between ST and TT, between ST and Dictionary and between TT and Dictionary. The distribution of these saccades during translation was calculated and depicted in Figure 17 and 18. The majority of switches were found between ST and TT, 75.6% and 72.9% in respective texts. This was predictable because the two interest areas were considered the main components of translation process and have received significantly longer dwell time comparing with the

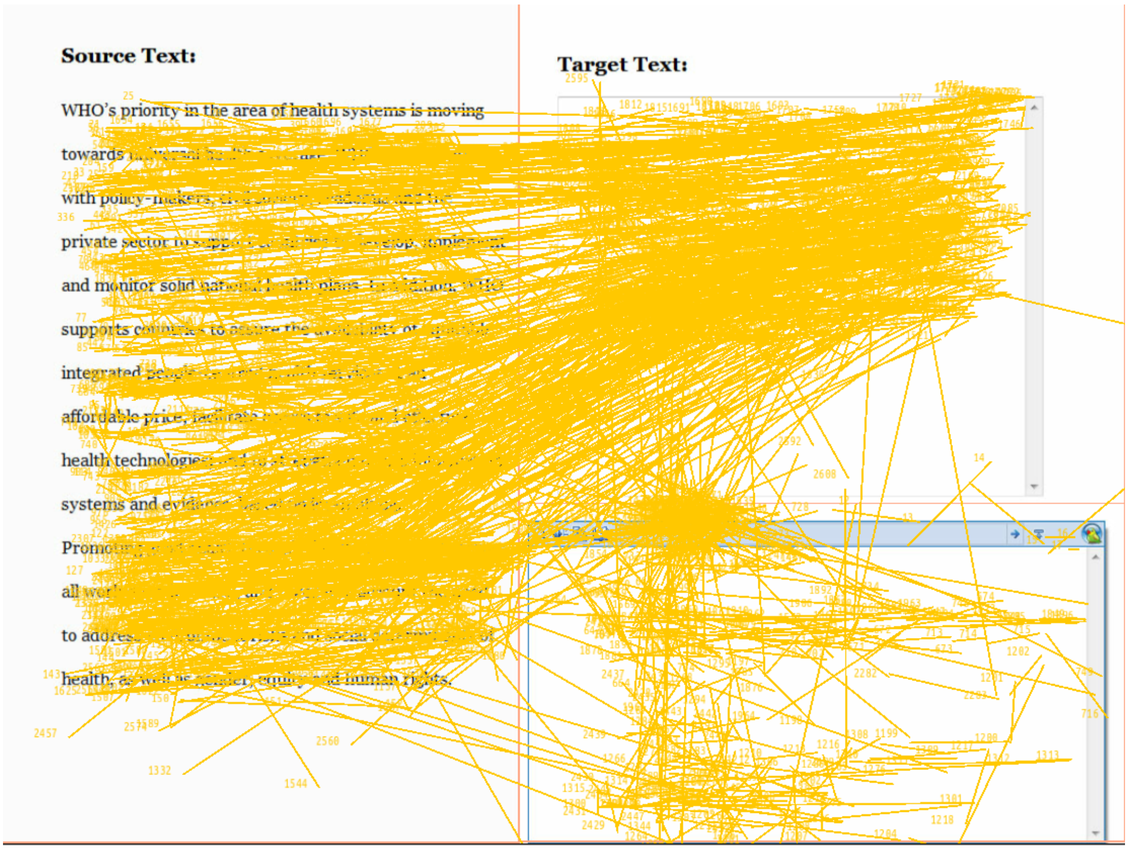

Figure 17 Visualization of saccadic movements across AOIs.

dictionary area. Regarding the shifts to or from dictionary, higher percentage was found in TT-Dictionary than ST-Dictionary saccades. It appears to resonate with Carl & Dragsted (2012)'s discussion on the monitor model, as mentioned in Section 1, indicating that ST understanding is prompted by problems occurring in TT production rather than by a lack of ST comprehension.

Two-way ANOVA was conducted to investigate the statistical significance between groups. It turned out to have main effect in directions of saccades ($F=72.983$, $p<0.001$) while no significance was found in Text Type ($F=0.440$, $p=0.510$) and the interaction ($f=0.650$, $p=0.526$). In addition, post-hoc analyses showed pairwise significance in ST-TT and ST-Dictionary ($p<0.001$) as well as ST-TT and TT-Dictionary ($p<0.001$). No significance was found in the pairs of ST-Dictionary and TT-Dictionary ($p=0.681$). The result confirmed that most of the shifts by translators were performed between ST and TT. On the other hand, attentional shifts

		Cross AOIs saccades			
		ST-TT	ST-Dict	TT-Dict	Total
Text 1	Count	1210	158	232	1600
	Percentage	75.6%	9.9%	14.5%	100.0%
Text 2	Count	1044	170	218	1432
	Percentage	72.9%	11.9%	15.2%	100.0%
Total	Count	2254	328	450	3032
	Percentage	74.3%	10.8%	14.8%	100.0%

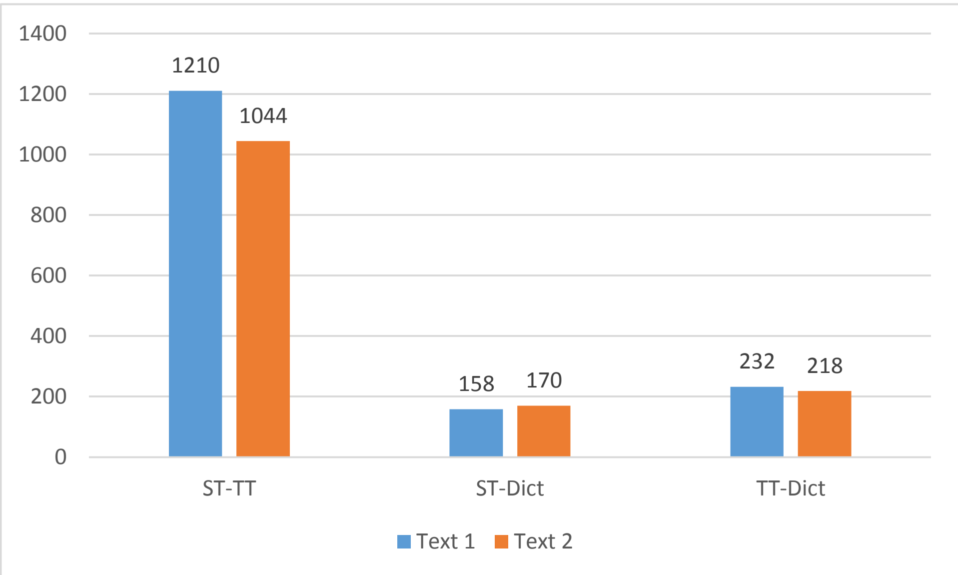

Figure 18 Counts and distributional percentage of cross AOIs saccadic movements (upper) and corresponding bar chart, provided for clarity (lower).

to dictionary from either ST or TT were not significant, and accordingly, the statistics defied the assumption suggested in the monitor model. The subsequent analysis was carried out to further examine the patterns of dictionary lookup activities.

4.6.2 Dictionary Reading Pattern by Saccades

As illustrated in the previous paragraph, saccades to dictionary area originated from either ST or TT area, and accordingly, it could be assumed that there were two kinds of dictionary lookup activities, which were, on one hand, prompted by problems of comprehension, or on the other hand, by the need during TT production. The fixation duration after cross AOIs saccades were analyzed to examine whether different patterns of dictionary lookup activity can be found statistically.

The descriptive statistics got in line with the assumption made by the monitor model that problems encountered during production required more cognitive attention from translators, displaying longer mean duration of fixations after TT-Dictionary saccades for both texts. Additionally, Text 2 received longer mean duration among both ST-Dictionary and TT-Dictionary saccades. No main effect was found from any of the three factors involved, including Text Type (F=0.680, p=0.416), the origin of cross AOIs movements (F=0.889, p=0.353) and the interaction (F=0.004, p=0.952). Furthermore, no significance was found in any t-test with in-group analysis, including analyses with saccade directions as independent factor (Text 1: F=0.210, p=0.653; Text 2: F=1.438, p=0.248), and those with text complexity as independent factor

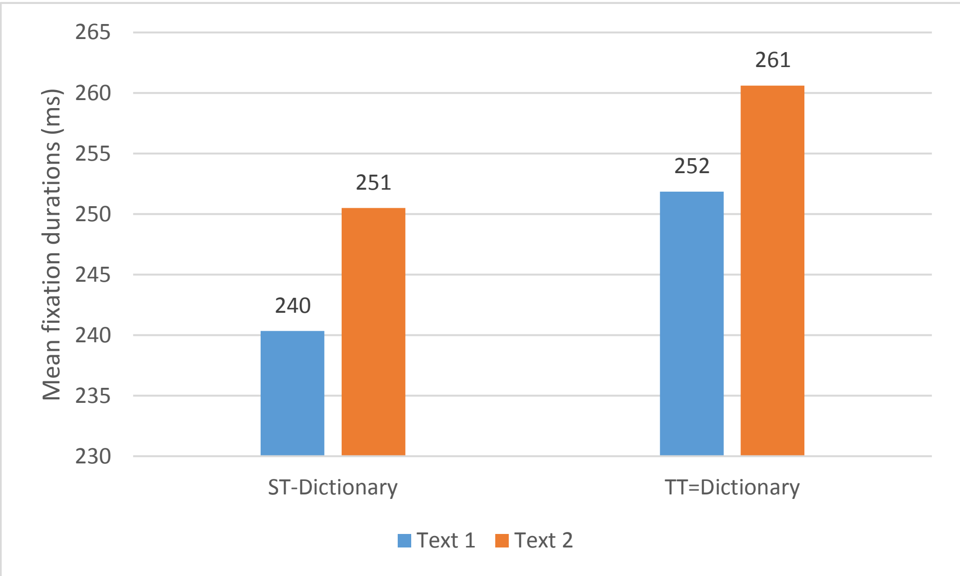

Figure 19 Mean duration of fixations in dictionary area after ST-Dictionary and TT-Dictionary saccades (ms).

(ST-Dictionary: F=0.975, p=0.338; TT-Dictionary: F=0.254, p=0.621).

The result failed to support the assumption, since no statistically significant patterns of dictionary lookup activities were found by analyzing the fixation duration after cross AOIs saccades. Despite the greater mean value from the TT-Dictionary group, modifications to the analysis should be taken to verify the assumption statistically. Few possible modifications can be considered. Firstly, the analysis by calculating the fixations after each cross AOIs saccade might be debatable. As implied in Jakobsen's micro-cycle model (discussed in Section 1), recurrent shifts between ST and TT could take place, and therefore, the rigid definition of fixations might not necessarily reflect to its previous saccade. For example, if a translator shifts from TT to ST to check the spelling and then shift to dictionary, the fixations, coded under ST-Dictionary group, might actually reflect a dictionary lookup activity for the need of TT production. Acknowledging this problem, a more delicate method to categorize the activities should be implemented.

In addition, the text complexity indicators adopted in this project might not be ideal for this analysis, since no detailed word-level study was involved. As illustrated in the descriptive statistics, Text 2, rated less complicated, actually gave rise to longer mean duration in both direction groups that Text 1. Even though the two texts were found to have similar percentages of low frequency words (explained in Section 3), it is possible that Text 2 involves a highly specialized vocabulary. For example, the word "orangutan" was looked up by most participants. In this regard, detailed word-level analysis might be necessary for the study of dictionary activities.

5. Conclusion

Quantitative data collected from eye tacking, possibly along with other modalities, has offered new insights into translation studies. According to Grucza (2013), eye tracking has been regarded as a “translation process research method,” and has been adopted to explore a wide range of questions concerning both translation and interpretation process. It has assisted researchers to harvest fruitful results on various aspects of translation, and provided quantitative explanations for various theoretical hypotheses. While researchers have adopted eye tracking to answer a diversity of questions in translation studies, this project returns to the fundamental questions about how cognitive attention of translators distributes and how the cognitive workload varies during translating.

In order to facilitate the discussion of concepts and the data, an overview of previous and ongoing researches in translation process studies, the mechanism of the technology and the theoretical interpretation of eye-tracking data were first presented in the project report. Afterwards, the empirical research of this project focused on mainly three aspects of translation process, including the general distribution of cognitive attention, cognitive workload of various subtasks and working style of translators. Accordingly, a few assumptions were raised and then confirmed or defied by the statistical results. The analyses resorted to various measures, including dwell time, fixation duration, pupil size and saccade counts, to investigate different questions, and moreover, the fixed factors such as level of text complexity as well as fixation and saccade grouping, either according to AOIs or subtasks of translation, were adopted in this project, to study how translators spend their cognitive efforts.

The first analysis was carried out on dwell time to examine the overall distribution of cognitive attention in the translation process, and predictably, TT production took up the highest percentage. The results indicated that both AOIs and Subtask Type would affect the distribution of cognitive attention, and text complexity rendered no significant effect in the overall distribution.

Subsequently, cognitive workload on respective AOIs and subtasks was investigated. Both fixation duration and pupil size were analyzed and discussed together. As a result, TT production and ST comprehension accounted for the greatest demand for cognitive workload than dictionary lookup and parallel attention. Further, text complexity had positive relationship with pupil size in subtasks including ST comprehension, TT production and dictionary lookup, implying the increase of cognitive workload along with complexity of processing. While discrepancy was identified in the results of fixation duration and pupil size, detailed subtask analyses were conducted. As a consequence, typing events would lead to longer duration regardless of the amount of cognitive workload required. The observation provided an avenue to identify more detailed subtasks, and additionally, it also offered an explanation to the drastic difference between fixation duration and pupil size in PA, which displayed long duration yet small pupil size. This unexpected result showing lower cognitive workload for PA was discussed and explained along with possible research questions.

The third part of this project strived to analyze the observable patterns of translators' working style by looking into saccadic movements. Cross AOIs saccadic movements and their probabilities were studied to investigated how translators shift their attention during translation. As predicted, shifts between ST and TT accounted for the greatest percentage. However, this analysis failed to explain the statistical significance between shifts to dictionary area from either ST or TT, which defied the assumption based on the monitor model that dictionary lookup activities were prompted by problems encountered during production. Further analysis was conducted on top of this assumption to look into different patterns of dictionary lookup activities in comprehension and production. The analysis studied the duration of fixations after cross AOIs saccades from either ST or TT to dictionary area. Nevertheless, even though the absolute mean value displayed observable longer duration of fixations after TT-Dictionary movements, but failed to achieve statistical significance. Discussions were then given to identify the difficulties of this analysis and possible directions were suggested for future research.

In this project, by means of several measures as cognitive workload indicators, typical translation subtasks were investigated in terms of their cognitive intensity, and further, generalized patterns of attentional distribution and working style were examined. The attempt to identify a list of translation subtasks as comprehensive as possible in this project could possibly help to identify the general patterns of translation behaviors or provide a more detailed picture of the micro-cycle of translation process.

The goal of this project is to examine the allocation of cognitive effort across different domains and subtasks of translation. With an aid from eye tracking, quantitative results have been obtained to reveal the cognitive intensity of translators when they were dealing with ST comprehension, TT production, dictionary lookup or even parallel processing. While this project has focused on certain factors, such as text complexity and its influence on cognitive workload, some other aspects concerning the translation process could also have been taken into consideration, for example, how time constraint or proficiency of translators would affect their working style and attentional distribution. Furthermore, if an objective and systematic method for evaluating TT output could be available, investigating the correlation between TT quality and the patterns of cognitive effort distribution would be very much beneficial to translation education and researches. Further studies of cognitive workload at the sentence or even the word level, for example, how translators react when they encounter odd or ungrammatical sentences, could also offer an interesting scope to the subtle cognitive variations during the whole translation process.

6. Appendix

6.1 Experimental Texts

6.1.1 Source Texts

(Text 1) Adapted from *World Health Organization*. Retrieved from: http://www.who.int/about/what-we-do/en/

WHO's priority in the area of health systems is moving towards universal health coverage. WHO works together with policy-makers, civil society, academia and the private sector to support countries to develop, implement and monitor solid national health plans. In addition, WHO supports countries to assure the availability of equitable integrated people-centred health services at an affordable price; facilitate access to safe and effective health technologies; and to strengthen health information systems and evidence-based policy-making. Promoting good health through the life-course cuts across all work done by WHO, and takes into account the need to address environment risks and social determinants of health, as well as gender, equity and human rights.

(Text 2) Adapted from *Greenpeace East Asia*. Retrieved from: http://www.greenpeace.org/eastasia/campaigns/forests/

The Earth cannot sustain life without healthy, thriving forests. They are home to over two-thirds of the world's species. They are like the green lungs of the planet, supplying us with oxygen and helping to balance rainfall and the climate. Yet our forests are at risk. They are being logged for chocolate, toothpaste, tissue paper, magazines, animal feed and more. They are being burned, degraded and logged at astonishing rates – as much as 80% of the world's forests are already destroyed. Unique wildlife like orangutans and the clouded leopard are pushed to the brink of extinction, while indigenous people are uprooted from their traditional homes.

6.1.2 Sample Target Texts by Random Participants

(Text 1)

世界衛生組織(WHO)目前在健康方面的首要任務是世界人口健康保障. 該組織與政策制定者, 公民社會, 學術界以及商界合作, 支持各國家發展, 推行及監測實際的國家人口健康方案. 另外, 世界衛生組織亦支持各國家, 確保能提供合理, 便宜, 又以人為本的醫療服務; 能夠使用安全有效的醫療科技; 能夠強化健康資訊系統, 以及注重現況的政策制定. 提高健康意識是世界衛生組織的工作總述, 組織亦負責發布影響健康的環境風險及社會因素, 還有性別, 平等與人權.

(Text 2)

沒有健康,有活力的森林, 地球便不能繁衍生命了. 森林是全球六成多物種的家園. 森林就好比地球的綠色肺部, 為我們提供氧氣, 有幫助我們調節雨水及氣候循環.

但我們的森林現正面臨危機, 人們為了巧克力, 牙膏, 廁紙, 雜誌,食用動物等等不斷地砍伐森林樹木. 樹木被燃燒,分解以及看法的速度遠超想像, 地球上多達八成的森林已經被破壞. 獨特的野生動物已經被推到滅絕的邊緣, 而森林的原住民們亦被驅趕離開他們的家園.

6.2 Participant Profile

	Sex	*Age*	*Cohort*	*Education*	*Native Language*	*Optical aid*
P1	F	22	2012	BA T&I, CityU	Mandarin	Yes
P2	F	23	2012	BA T&I, CityU	Cantonese	Yes
P3	F	22	2012	BA T&I, CityU	Cantonese	No
P4	F	20	2013	BA T&I, CityU	Mandarin	Yes
P5	F	22	2012	BA T&I, CityU	Mandarin	No
P6	F	23	2012	BA T&I, CityU	Cantonese	No
P7	F	22	2012	BA T&I, CityU	Mandarin	Yes
P8	F	22	2012	BA T&I, CityU	Cantonese	Yes
P9	F	22	2012	BA T&I, CityU	Cantonese	Yes
P10	F	23	2012	BA T&I, CityU	Cantonese	Yes
P11	F	22	2012	BA T&I, CityU	Cantonese	No
P12	F	22	2012	BA T&I, CityU	Cantonese	No
P13	F	23	2012	BA T&I, CityU	Cantonese	No
P14	F	23	2012	BA T&I, CityU	Mandarin	Yes

6.3 Table of Statistics

Table 1 Results from the two-way ANOVA of dwell time.

	Dwell Time by AOIs (ms)				
	SS	df	MS	F-stat	p-value
AOI	8.33192E+11	2	4.16596E+11	16.749	0.000
Text Type	33330007074	1	33330007074	1.340	0.252
*AOIs * Text Type*	31066882956	2	15533441478	0.625	0.539
Error	1.34314E+12	54	24872901918		
Total	5.89353E+12	60			

	Dwell Time by Subtask Type (ms)				
	SS	df	MS	F-stat	p-value
Subtask Type	1.42868E+12	3	24841768655	25.521	0.000
Text Type	24841768655	1	2.69932E+12	1.331	0.252
*Subtask Type * Text Type*	31019767224	3	4.76227E+11	0.554	0.647
Error	1.34354E+12	72	10339922408		
Total	5.52741E+12	80	18660326866		

Table 2 Results from the two-way ANOVA of the fixation duration and pupil size.

	Fixation duration by AOIs (ms)					Pupil size by AOIs (au)				
	SS	df	MS	F-stat	p-value	SS	df	MS	F-stat	p-value
AOIs	23978598.42	2	11989299.21	402.184	0.000	67643200.3	3	22547733.43	103.192	0.000
Text Type	15056.659	1	15056.659	0.505	0.477	47899866.76	1	47899866.76	219.219	0.000
AOIs * Text Type	175605.197	2	87802.598	2.945	0.053	6922050.354	3	2307350.118	10.560	0.000
Error	1746238416	58578	29810.482			12686460865	58061	218502.28		
Total	5511215676	58584				1.51299E+11	58069			

	Fixation duration by Subtask Type (ms)					Pupil size by Subtask Type (ms)				
	SS	df	MS	F-stat	p-value	SS	df	MS	F-stat	p-value
Subtask Type	27026342.2	3	9008780.733	300.213	0.000	67643200.3	3	22547733.43	103.192	0.000
Text Type	110482.987	1	110482.987	3.682	0.055	47899866.76	1	47899866.76	219.219	0.000
Subtask Type * Text Type	234220.049	3	78073.35	2.602	0.050	6922050.354	3	2307350.118	10.560	0.000
Error	1742290034	58061	30007.923			12686460865	58061	218502.28		
Total	5488480295	58069				1.51299E+11	58069			

Table 3 Results from the two-way ANOVA of shift probabilities and fixation durations in dictionary after cross AOIs saccades.

	Shift probabilities (counts)				
	SS	df	MS	F-stat	p-value
Cross AOIs	474285.733	2	237142.867	72.983	0.000
Text Type	1430.817	1	1430.817	0.440	0.510
*Cross AOIs * Text Type*	4224.933	2	2112.467	0.650	0.526
Error	175462.7	54	3249.309		
Total	1281073	60			

	Mean durations in dictionary after cross AOIs saccades (ms)				
	SS	df	MS	F-stat	p-value
Cross AOIs	1050.927	1	1050.927	0.889	0.353
Text Type	804.155	1	804.155	0.68	0.416
*Cross AOIs * Text Type*	4.299	1	4.299	0.004	0.952
Error	37823.915	32	1181.997		
Total	2304704.549	36			

Reference

1. Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more. *Journal of Eye Movement Research*, 3(3), 6, 1-12.

2. Ball, L. and Poole, A. (2006). Eye tracking in HCI and usability research. In C. Ghaoui, *Encyclopedia of Human Computer Interaction* (1st ed., pp. 201-209). Hershey: Idea Group Publishing.

3. Bell, R. (1991). *Translation and the Nature of Philosophy: A New Theory of Words.* London and New York: Routledge.

4. Bergstrom, J. & Schall, A. (2014). *Eye tracking in User Experience Design*. Waltham: Elsevier Inc.

5. Beymer, D., Russell, D., & Orton, P. (2008). An eye tracking study of how font size and type influence online reading. *Paper presented at the Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction,* 2. Liverpool: United Kingdom.

6. Boehm, D.C. (1993). Mozartians, Beethovians, and the teaching of writing. *The Quarterly*, 15(2), 15-18.

7. Bojko, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. Brooklyn, New York: Rosenfeld Media.

8. Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, 42, 361-377.

9. Bruneau, D., Sasse, M. A., & McCarthy, J. D. (2002). The eyes never lie: the use of eye tracking data in HCI research. In *Proceedings of the CHI'02 Workshop on Physiological Computing*. NY: ACM Press.

10. Carl, M., Dragsted, B. & Jakobsen, L. (2011). A taxonomy of human translation styles. *Translation Journal*, Vol. 16, No. 2

11. Carl, M., & Dragsted, B. (2012). Inside the monitor model: processes of default and challenged translation production. *TC3, Translation: Computation, Corpora, Cognition*, 2(1), 127-145.

12. Catford, J. (1965/ 2000) *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press.

13. Cowen, L., Ball, L., & Delin, J. (2002). An Eye Movement Analysis of Web Page Usability. In X. Faulkner, J. Finlay, & F. Détienne (Eds.), *People and Computers XVI - Memorable yet Invisible: Proceedings of HCI 2002* (pp. 317-335). London: Springer London.

14. Danks, J. & Griffin, J. (1997). Reading and translation, A psycholinguistic perspective. In J. Danks, G. Shreve, S. Fountain & M. McBeath (Eds.), *Cognitive Processes in Translation and Interpreting* (pp.161-175). London: Thousand Oaks, Calif., Sage Publications, Inc.

15. Danks, J. H., & Griffin, J. (1997). Reading and translation: A psycholinguistic perspective. In J. H. Danks, G. M. Shreve, S. B. Fountain & M. K. McBeath (Eds.), *Cognitive Processes in Translation and Interpreting*, 161-175. Thousand Oaks: Sage.

16. Dragsted, B. & Carl, M. (2013). Towards a classification of translator profiles based on eye-tracking and keylogging data. *Journal of Writing Research*, 5(1), 133-158. http://dx.doi.org/10.17239/jowr-2013.05.01.6

17. Doherty, S., O'Brien, S., & Carl, M. (2010). Eye tracking as an MT evaluation technique. *Machine Translation, 24*(1), 1-13.

18. Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455-470. doi:10.3758/bf03195475.

19. Duchowski, A. (2007). *Eye Tracking Methodology*. London: Springer.

20. Ehrensberger-Dow, M. & Massey, G. (2013). Indicators of translation competence: Translators' self-concepts and the translation of titles. *Journal of Writing Research*, 5(1), 103-131.

21. Gile, D. (1995). Basic concepts and models for interpreter and translator training. *Benjamins Translation Library*. http://dx.doi.org/10.1075/btl.8(1st).

22. Gile, D. (2011). Errors, omissions and infelicities in broadcast interpreting. *Methods and Strategies of Process Research*, 201-218. http://dx.doi.org/ 10.1075/btl.94.15gil

23. Goldberg, H. J., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye:* Cognitive *and Applied Aspects of eye Movement Research,* 493-516. Amsterdam: Elsevier.

24. Grucza S. (2013). The investigation of translation competences and translation processes. *Вестник Московского государственного лингвистического университета*, 4(664), 51-60.

25. Hess, E. H. & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, 132(3423), 349-350.

26. Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, UK: Oxford University Press.

27. Hvelplund, K. T. (2011). *Allocation of Cognitive Resources in Translation: An Eye-Tracking And Key-Logging Study*. PhD thesis, Copenhagen Business School.

28. Jääskeläinen, R. (2000). Focus on methodology in think-aloud studies in translating. In S. Tirkkonen-Condit & R. Jääskeläinen, *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research* (1st ed.). Amsterdam: John Benjamins Publishing.

29. Jääskeläinen, R. (2011). Back to basics: designing a study to determine the validity and reliability of verbal report data on translation process. In S. O'Brien (ed.), *Cognitive Explorations of Translation* (1st ed., pp. 15-27). London: Continuum International Publishing Group.

30. Jacob, K. & Karn, S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises, *Mind*, 2(3): 4.

31. Jakobsen, A. (2003). Effects of think aloud on translation speed, revision, and segmentation. *Perspectives in Process Oriented Research*, 69-95. http://dx.doi.org/10.1075/btl.45.08jak.

32. Jakobsen, A.L. (2011). Tracking translators' keystrokes and eye movements with Translog. In C. Alvstad, A. Hild & E. Tiselius (Eds.), *Methods and Strategies of Process Research*, 37-55. Amsterdam: John Benjamins. doi: 10.1075/btl.94.06jak

33. Jakobsen, A. & Jensen, K. T. (2008). Eye movement behaviour across four different types of reading task. *Copenhagen studies in language*, 36, 103-124.

34. Jensen, K. T. (2009). Indicators of Text Complexity. *Copenhagen Studies in Language*, 37, 61-80.

35. Jensen, K. T. (2011). Distribution of attention between source text and target text during translation. In S. O'Brien, *Cognitive Explorations of Translation* (1st ed., pp. 215-236). London: Continuum International Publishing Group.

36. Just, M. A. & Carpenter, P. A. (1976). Eye fixations and cognitive processes, *Cognitive Psychology,* 8, 441-480.

37. Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, 87(4), 329-354. http://dx.doi.org/10.1037/0033-295x.87.4.329.

38. Krings, H. P. (1986b). Translation problems and translation strategies of advanced German learners of French (L2). In J. House, & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition Studies*, 263-276. Tübingen: Gunter Narr.

39. Munday, J. (2001). *Introducing Translation Studies*. London: Routledge.

40. Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45(1), 272-288.

41. O'Brien, S. (2010). *Cognitive Explorations of Translation*. New York: Continuum International Publishing Group.

42. Poole, A., and Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research. In C. Ghaoui (ed.), *Encyclopedia of human computer interaction. Idea Group, Pennsylvania*, 2005, 211-219.

43. Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160-174. http://dx.doi.org/10.1037/0096-3445.109.2.160.

44. Radach, R., Kennedy, A. & Rayner, K. (2004). *Eye Movements and Information Processing during Reading*. Psychology Press. ISSN/ISBN:184169956x hbk

45. Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422. http://dx.doi.org/10.1037/0033-2909.124.3.372.

46. Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice Hall.

47. Tirkkonen-Condit, S. (2005). The monitor model revisited: evidence from process research. *Meta*, *50*(2), 405. http://dx.doi.org/10.7202/010990ar

48. Tycová, B. (2015). *Eye-tracking in Translation Process Research* (Master's Diploma Thesis). Department of English and American Studies, Masaryk University.

49. Vinay, J. & Darbelnet, J. (1958/ 1995). *Comparative Stylistics of French and English*. Benjamins Translation Library. http://dx.doi.org/10.1075/btl.11

50. White, J. (2003). How to Evaluate Machine Translation? A Translator's Guide, pp.211-244.

51. Zwierzchoń-Grabowska, E. (2011). Okulograficzne wsparcie badań nad tłumaczeniem pisemnym. In: Lingwistyka Stosowana/ Applied Linguistics/ Angewandte Linguistik. 4. 199–210.