國立臺灣師範大學翻譯研究所
博士論文
Doctoral Dissertation
Graduate Institute of Translation and Interpretation
National Taiwan Normal University

訓練與經驗對英譯中視譯認知歷程影響之研究：

眼動與產出之整合分析

An Integrated Eye-tracking Study into the Cognitive Process of English-Chinese Sight Translation: Impacts of Training and Experience

指導教授：陳子瑋博士、蔡介立博士
研究生：何承恩
中華民國一〇六年三月

Advisors
Dr. Chen, Tze-Wei
Dr. Tsai, Jie-Li
By
Ho, Chen-En (Ted)
March 2017

摘要

視譯長久以來僅被當成課堂學習的輔助工具。然而，近年視譯於實務中的使用機率已有所提升，社區口譯等子領域甚至更將視譯列為必備技能，加上多項研究已證明，視譯確有不同於其它口譯模式的難處，因此視譯實值得深入研究。由於眼動儀可觀測閱讀情形，因此以此儀器研究視譯，可直接了解視譯的認知歷程有助口譯研究的學科繼續發展。

本研究分析 17 位資深口譯員、18 位口譯學生、18 位未受訓的雙語使用者，期能了解閱讀目的是否會影響閱讀行為，找出訓練與口譯實務經驗造成的影響，並細部觀察視譯過程中的「前導閱讀」（reading ahead）與停頓行為。結果發現，閱讀目的不同，的確會改變閱讀行為；然而，不同作業之間的閱讀行為仍有部分相似。閱讀第一遍（first pass）時，為了理解而閱讀（默讀）與為了翻譯而閱讀（視譯）的行為非常相似，到了第二遍（second pass）之後才開始出現差異。朗讀在閱讀第一遍時耗費的時間遠多於其它兩者，但進入第二遍之後，與默讀的相似度便大幅提升。整體而言，默讀與朗讀的認知負荷較接近，後者略高，視譯則最為費力。

訓練與口譯經驗究竟有何影響？本研究發現，口譯專家的視譯品質最高，再者為口譯學生，最低者為未受訓的雙語者。整體數據（global data）顯示，經過訓練者，視譯總時間與總凝視數量皆顯著少於一般雙語者。不過，譯文總字數、細部閱讀指標（local reading indices）的平均凝視時間，以及閱讀廣度（reading span）皆無組間差異。由此推知，差異應主要來自各組的視譯過程。經過訓練者，畫面一出現，短暫默讀後便會開始視譯，未受訓者絕大多數則會完整將內容讀過，因此一開始的停頓顯著較長。此外，經過訓練，每次開口前的凝視次數也顯著較少，口譯專家更是極端，就算遇到中英語言明顯不同的分枝結構（principal branching direction units），也幾乎不會受阻而出現較長停頓。口譯學生的做法與專家類似，只是轉換技巧較不熟練，偶爾停頓較長、凝視次數較多；相對而言，未受訓的雙語者出現長時間停頓與多次凝視的頻率較高。整體而言，經過訓練，可察覺的停頓數量便會減少，流暢度較高，產出速度也較快。至此，訓練的影響已相當明顯，品質與速度皆會顯著提升。另一方面，經驗影響的似乎主要是正確度，以及開始視譯後的前導閱讀與停頓行為。

累積了足夠的經驗之後，專家的遲疑停頓（hesitation pause）與結構停頓（juncture pause）的比例已無顯著差異，其他兩組的遲疑停頓則顯著較多，但三組停頓時，凝視分枝結構的機率都在一半以上。最後，受過訓練者，包括資深口譯員與口譯學生，不管是哪種任務，處理中英分枝結構相異處所耗費的時間與非

相異結構並無顯著差異，至於一般雙語者，於默讀任務時已可觀察到，在第一遍閱讀中比較晚期的指標與非首遍閱讀的訊息統整階段，分枝結構相異處顯著較費心力，進入視譯，除了前述兩者之外，分枝結構相異處的單字平均凝視總時間也顯著較長，顯示未受訓者於處理語言分枝結構單位方面較為吃力。

關鍵詞：視譯、眼動、認知歷程、前導閱讀、訓練與經驗

Abstract

Sight translation has long been regarded as nothing more than a pedagogical tool. However, the role of sight translation has become more important in practice, and even necessary for certain fields such as community interpreting. What's more, with a growing number of studies proving this task to be no easier than other modes of interpreting, we are now granted an opportunity to further develop the discipline of interpreting studies by looking into the cognitive process of sight translation, which can be directly observed with an eye tracker.

This study analyzed data of 17 experienced interpreters, 18 interpreting students, and 18 untrained bilinguals, hoping to 1) find out if reading purpose affects reading behavior in different ways, 2) understand the impacts of training and experience, and 3) look more closely at the behavior of reading ahead and pausing during sight translation. The results showed that, different reading purposes did change reading behavior, but there were still similarities between tasks. Reading for comprehension was similar to reading for sight translation in the first pass of reading, while the two started to diverge in the second pass. On the other hand, reading aloud, a task requiring more efforts in the first pass, began to resemble silent reading more than sight translation in the second pass. Generally speaking, the cognitive load imposed on silent reading was similar to (and a little less than) reading aloud, and sight translation was significantly more strenuous.

Turning to the impacts of training and experience, it was found that the quality of sight translation was the highest for experts, followed by trainees and then untrained bilinguals. Global data showed less total time and fewer fixations for trained participants than bilinguals. Nevertheless, word counts, mean fixation duration in all local reading indices, and even reading span failed to show any difference. As it turned out, the difference lay in how each group proceeded with the task. Trained participants started sight-translating the text shortly after each trial began, with few fixations of reading ahead, when bilinguals mostly read through the text first, leaving a rather long silence at the outset. In addition, participants with training had significantly fewer fixations before uttering each Chinese character. Experts were extreme in that they rarely got bogged down, even when encountering contrastive linguistic structures. Trainees also manifested a similar tendency, though not as adept at the reformulation skills as experts. Bilinguals, on the opposite, showed much more fixations and longer pauses from time to time. Overall, observable pauses were fewer for those with training, and average verbal gaps were shorter as well, leading to higher fluency and a quicker pace. While the impacts of training were obviously on quality and speed, experience seemed to mainly affect accuracy and the behavior of reading ahead and pausing once sight translation began.

Experts, with ample experience, had a non-significant gap between the percentage of hesitation pause and that of juncture pause, whereas the other two groups had significantly more hesitation pauses. Notwithstanding, all groups tended to fixate on principal branching direction (PBD) units at least 50% of the time during pauses. Finally, trained participants (i.e., both experts and interpreting students) did not show any difference between the time spent on processing PBD and non-PBD units in all three tasks. On the other hand, untrained bilinguals already inclined to spend significantly longer time on PBD units during silent reading in a relatively later stage of first-pass processing and for meaning integration in non-first-pass reading. When performing sight translation, the results during silent reading were replicated; what was more, total viewing time on PBD units was also significantly longer, indicating that processing PBD units were more effortful for untrained bilinguals.

Keywords: Sight translation, eye movement, cognitive process, reading ahead, training and experience

致謝

沒想到動筆撰寫致謝的這一天默默地就到了。踏進師大翻譯所，從英語教學轉入了口筆譯的世界。這個地方別有洞天，環顧四周，一位比一位優秀，回頭看看自己，卻覺得茫茫然，不知道自己的定位，也不知道該往何處去。打滾掙扎了一陣，終於確立了攻讀博士這條路。這條路走得孤單，因為念口譯的人很少這麼選擇；這條路走得艱辛，因為許多事情都需要重新摸索、重新適應。好像每每下了決定，當下自覺放了自己一條生路，人生嘛，何必？但走到一半常不住覺得自己是往滿布荊棘的那條小徑闖，遇過的問題、心魔，避得了一時，避不了一世。所幸，一路上不斷遇到貴人，跌跌撞撞個幾年，也總算走過了那昏暗的坑道，看到另一頭的光。

對父母的感謝絕對少不了，生我育我，尊重我的選擇、不斷給予各方面的支援，還要提供建議、輔導，這樣的恩情說比天高肉麻了，但扎扎實實，點滴在心頭。感謝兩位指導教授各方面的協助、指引，少了任一位，這條路都無法順利走完。子瑋老師從口譯實務到研究，甚至是教學，樣樣都手把手傳授，從不吝於花時間耐心解釋與討論，最終會選擇認知科學做為探究口譯本質的途徑，也是老師指引的一條明路，開了眼界，初窺了非常有趣的領域。人常說，做研究是孤獨的，但選了這條路，發現新知的喜悅，值了。介立老師更一手促成這篇博論的誕生。從閱讀心理學與眼動儀相關知識，到程式語言、實驗設計、資料蒐集、結果詮釋，樣樣都是經過老師的大力支援與照看，才關關難過關關過。常常帶著滿肚子的疑惑去叨擾老師，每每得到耐心的解釋，茅塞頓開。摸索過程中，更發現程式語言的有趣之處。說老師點亮了明燈、照亮了另一條路也不為過，雖然不時撞得滿頭包，但箇中趣味，實在難以言喻。

感謝論文口試的所有委員，提出的各個問題與建議都切中要害，回饋不僅全面，更是細節俱到，許多都是撰寫論文時的盲點，期待自己未來也能達到同樣的境界。感謝所有參與實驗的口譯專家、口譯學生、未受訓者。到政大做實驗真的需要耗費許多寶貴時間，這篇論文應該獻給所有的受試者。感謝一路上替我加油打氣的老師、學長姐、學弟妹與親朋好友，三不五時的垃圾話（咦？）是讓人繼續走下去的動力。最後，要感謝生命中最重要的尼姆！寫論文時只專心做一件事的癡心妄想從未實現過，教學、研究計畫、口筆譯工作夾雜，寫論文也不時遇到磨難。身心俱疲，但卻難以入眠，亦或是時常想到論文而驚醒。好幾次都興起放棄的念頭，是尼姆陪我走過每一段路，連假日都陪我趕工，不時給予安慰、鼓勵。學校生涯終於告一段落了！新的開始，我會帶著滿滿的感謝，闖他一闖。

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

Interpreting has been a common practice that began long ago, but it had seldom come to researchers' attention until 20th century. As the needs for international conferences or multilingual meetings among heads of nations or foreign affairs representatives rise, or due to the necessity for international tribunals to be held (specifically, the Nuremberg Trial as an example), interpreters per se and interpreting as a professional activity have become subjects of scientific investigation (Pöchhacker, 2015). It is also because of the skyrocketing needs of interpreters, training programs abound accordingly. Though the notion of professionalization started to gain ground some time ago, and interpreting did attract some scholarly interest in other disciplines, this mode of communication was mostly a subtopic that was within the research scope solely of professional practitioners and a small fraction of academics in the world (Pöchhacker, 2004).

The early 1990s is a historic turning point. The field of interpreting has since successfully established itself as an independent discipline, and much more connections with other academic disciplines have been bridged. The realm of interpreting research regarding topics, scopes, and methodology has also since diversified, taking this field of study to an international level (Pöchhacker, 2004). As a result, conferences on interpreting research, postgraduate-level theses, and journal papers multiplied due to wide recognition from the general public and academia (Gile, 2001). It was also at this time that more doctorate programs came on board to instill new momentum into interpreting research (Moser-Mercer & Setton, 2000).

In the beginning stage of academization, studies on conference interpreting was the mainstream, due to the “halos” bestowed onto interpreters that showed up on the stage or at the venues of high-end conferences or meetings. Interest was directed towards how one could adequately recall messages conveyed by speakers or perform such work as listening and speaking at the same time, and, of course, how one can be trained to become a professional practitioner. For example, Shlesinger (2000; cited from Mizuno, 2005) examined the working memory (WM) model proposed by Baddeley (1992) and found that concurrent activities of listening and speaking does exert a high cognitive load and is detrimental to interpreters’ retention of information in the working memory. Gile (1997) explicated the cognitive mechanisms of simultaneous interpreting, consecutive interpreting, and sight translation, and also how limited cognitive resources may affect one’s performance. Studies at that time were mainly curious about the skill per se.

Later on, because of large-scale immigration and the rising needs for sign language communication, community interpreting became a prevalent activity; issues that would typically concern community interpreters were increasingly added to the list of research topics in researchers’ mind, including settings and context, the role and identity of the interpreter, the power relations among interlocutors, or ethics and skillsets required in each unique context, to name just a few. In addition, due to the fact that qualified professional interpreters with suitable language combinations are not always readily available when one is needed in the legal, medical or social service context, some issues regarding expertise or the use of “natural” interpreters have especially piqued the interest of scholars (Pöchhacker, 2008, 2015). The trend of a "social turn" in research topics is further corroborated by Liu (2011) with an examination of papers published in *Interpreting* between 2004 and 2009 (see also Pöchhacker, 2008).

Almost a century after the beginning of interpreting studies, the broadening of the research repertoire is clearly a fact: from the skills and strategies, theories of interpreting, training of interpreters, and performance to work environment, communicative settings, roles and ethics, work-related technology, and the whole industry, including client relationships and professionalization (see also Ho, 2013; Liao, 2007). However, sight translation has rarely showed up as a topic of interest. Part of the result may come from the popular belief that this skill, as a hybrid of translation and interpreting, has less practical utility in the daily tasks of a conference interpreter, for whom consecutive and simultaneous interpreting take up almost the entire scope of work. Even if there are written materials prepared for participants' reference, the interpreter can always let go of the document at hand and focus solely on the main medium of message transmission: voice of the speaker. Considering that sight translation can be used to help students develop the skill of parsing or chunking, which may help ease the heavy cognitive burden of stacking up too much information before reformulation in simultaneous interpreting, many interpreter trainers tend to use sight translation as a pedagogical tool in class at best (Agrifoglio, 2004; Viaggio, 1995; Weber, 1990).

Nonetheless, in the world of community interpreting, sight translation is a task quite regularly performed. All too often, documents including briefs, affidavits, statements, or exhibits are presented in court without interpreters given copies of the documents in advance. In this setting, where average citizens' basic human rights are at stake, a competent sight translation skill is a prerequisite for all content to be translated into another language faithfully without any addition, deduction, or manipulation (Sampaio, 2007); National Association of Judiciary Interpreters and Translators and even International Organization for Standardization have both officially recognized sight translation as a necessary skill for court interpreters (Erickson, Bonet, Festinger, Framer, & Macfarlane, 2006; ISO 13611:2014). The same emphasis on the importance of sight translation is also shared by institutions working on medical interpreting, such as National Council on Interpreting in Health Care (2009) and International Medical Interpreters Association (2009).

Interestingly, even conference interpreting scholars have started to recognize an ever-bigger role played by sight translation skills. For example, Weber (1990) pointed out that sight translation could be used as a tool to prepare oneself for an incoming conference by actually speaking out loud technical terms, lingo, or specific ways of expression preferred by professionals in a certain field to build muscle memory. In addition, sight translation also has its place in consecutive and simultaneous interpreting, for one may at times run into the need to render a manuscript or PowerPoint slides prepared by the speaker. Some scholars eyeing the translation industry also observe that sight translation, coupled with speech recognition or audio recording software, may be a new method to boost translators' productivity and efficiency so that they can meet diverse requirements coming from clients while maintaining acceptable quality (see Biela-Woùoñciej, 2007; Dragsted & Hansen, 2009; and Dragsted, Mees & Hansen, 2011). With all its significance in the daily practice of interpreters or even translators, sight translation has every reason to be more than a pedagogical tool and deserves further exploration.

Previous studies on sight translation tried to understand the nature of this mode of interpreting and proved that this skill is just as unique and difficult as simultaneous and consecutive interpreting. Some focused on the cognitive load in the process, while some others paid attention to error types that are prevalent in each interpreting mode. Still others contemplated on how interpreters can smoothly sight-translate a text as if they were simply reading a text out loud in another language. The findings were fruitful, but some flaws in the methodology were quite common in these studies, including a rather small pool of participants and unclear standards for grouping, a fixed sequence for tasks and stimuli, a lack of control over the materials that led to practice effects and incomparable results, and also inferences that could not be backed by the empirical evidence at hand.

It is also a pity that most papers were product-based analyses, which gave us the end results but could not help us unveil the mysterious cognitive process of sight translation. Viezzi (1989) proved that content retention rate was lower in sight translation than in simultaneous interpreting, especially when the syntactic structures were widely diverse between the languages required by the task. Her (1997) and

Chang (2008) repeatedly found syntactic structures to hamper interpreters' smooth rendition. Agrifoglio (2004) and Chiang, Kuo and Chen (2009) each observed a high percentage of expression errors and inadequate pauses in the output of sight translation. Through output analysis, we have seen firsthand the difficulties of sight translation, which are no less challenging than consecutive or simultaneous interpreting. However, what factors impact sight translation production? What leads to the faltering or stammering rendition, at which point, and how? How exactly is sight translation done? These questions can only be solved when the above findings are coupled with a look in the process, that is, how information is received, deciphered, reformulated, and produced.

To disclose the cognitive process of sight translation, an examination of reading has to be included, since it is the only channel through which we gather information. Forerunners in the field of cognitive psychology have elucidated that reading itself is a complex mechanism which involves phonological and morphological decoding, meaning extraction, and even constant mapping of meaning on the word level to that which is conveyed through syntactic structures to help the reader get the full picture. In bilingual reading, factors such as the readers' linguistic competences in each language, the language environment, or even the knowledge and features of the second set of language systems further play a role and lead to facilitative or inhibitory effects on information retrieval. Specifically, syntactic differences in two languages have clear impacts on not just bilingual reading (Fernández, 2002; Frenck-Mestre, 2002), but also sight translation (Chang, 2008; Her, 1997; Viezzi, 1989), a mode of interpreting that requires the interpreter to activate both language systems but at the same time suppress structural disparities so as to ensure correct comprehension in one language and natural production in another in rapid succession. Hence, distinctive syntactic structures serve as an important reason for the present study's examination of English-Chinese sight translation.

We have already learned much about reading from the field of cognitive psychology and have also uncovered some interesting truths about interpreting. Yet, these two worlds seem to flourish in parallel without much connection. Sight translation is the skill that links reading and interpreting. Hence, a closer look at how reading is utilized during a mode of interpreting is a contribution of this study. To give a general picture of how sight translation is conducted and what really happens in the process, and to answer a few questions frequently asked by trainers and scholars, this study aimed to outline the cognitive process by tracking how interpreters read the texts to be translated via an eye tracker. Although four M.A. theses in Taiwan, including Huang (2011), Chen (2013), Su (2013), and Hsieh (2014), have already offered some valuable observation, all of their topics were confined to Chinese-to-English sight translation. This dissertation intended to complement previous findings with an opposite direction, which has never been investigated whatsoever, to lay the foundation for a deeper understanding of sight translation.

With features of the reading process as the backbone of this study, supplemented with quality assessment, we set out to tackle three main issues that frequented topics of papers in relevant fields, including the influence of reading purpose, training, and interpreting experience. We first tried to identify whether reading purposes would affect reading behavior. Then, we looked into how training would further influence reading behavior during sight translation and its output. Lastly, we saw how interpreting experience would make interpreters adjust the way they read and sight translate texts and what impacts experience would have on the oral rendition.

Chapter 2 Literature Review

This chapter was divided into a few sections. The first section defined what sight translation is and what cognitive process or skills are involved when performing this mode of interpreting. It goes without saying that an introduction of reading as a complex activity is necessary, as is bilingual reading. The characteristics of bilingual reading are closely relevant to this study because all interpreters (and sight-translators) are users of at least two languages. The last part of this section entailed why eye tracking would be the best fit for this study and also the eye-movement indices frequently used, and some brief explanation of the importance of those indices and their role in revealing the cognitive process of reading and sight translation. In addition, general linguistic structures of Chinese and English were compared and contrasted.

Next, we discussed findings and limitations of "traditional" research on sight translation, and also some other more non-traditional studies on the same topic. The section that followed would address recent eye tracking studies on Chinese-English sight translation. Lastly, the discussion turned to what remains to be done if we wish to more understand sight translation as a sophisticated activity and the grounds for this current dissertation.

2.1 Definition of sight translation

Sight translation, by definition, is a special type of simultaneous interpreting in which written texts are rendered orally (Pöchhacker, 2004). In light of this fact, some also refer to sight translation as a “hybrid” form of interpreting and translation, since it resembles the former in its way of conveying messages and the latter in that the reception of incoming content is through reading (Chen, 2015).

Even until today, sight translation has received relatively little attention precisely because of its hybrid feature. Translators enjoy unlimited time to ponder over the linguistic differences between their working languages and the final product, theoretically, so that they don’t see the need to apply sight translation to their task at hand. This is even truer for interpreters, whose work always consists of interpreting messages of speakers, not writers. At the end, sight translation is treated at best as pedagogical tools in the classroom (Weber, 1990). Trainees learn sight translation skills to chunk messages into smaller units so that they don’t get bogged down by a huge amount of incoming information during simultaneous interpreting when language structures are distinct in their working language pair. Some trainers also ask students to apply this kind of skill just to explore possibilities for dealing with sentences with an intricate structure.

However, in addition to the utility aspect we mentioned in the first chapter, sight translation itself is also intriguing from a cognitive perspective. Two different channels are invoked to complete the task of sight translation, including one’s reading ability to properly receive messages and speaking ability to reproduce content in another language. We can reasonably expect that reading as a sophisticated activity and cross-language differences that are more prominent in writing will have their impacts on the successfulness of a sight-translator’s rendition.

2.2 Reading

2.2.1 Reading in general

Reading is an inevitable part in sight translation, as it’s the only source of input. This task involves the procedure of extracting information from written words and constructing a mental representation of the whole text or comprehending the meaning conveyed. Within this process that is cumbersome in nature but rather quick in relation to time, a few fundamental subtasks have to be completed before we can determine whether the reading task is done successfully.

First of all, word identification is a core element and a prerequisite of reading that we encounter early on. Information is purely embedded in written symbols, and therefore orthographic decoding is the first and foremost step (de Groot, 2013), among which letter recognition plays the central role. Individual features such as horizontal or oblique lines are preserved in readers’ mental storage, and the number of matching features decides whether a certain letter will be activated or not (Rayner, Pollatsek, Ashby, & Clifton Jr., 2012). Under this efficient framework, our brain can save a huge amount of cognitive resources and quickly identify letters to trigger meaning access in later stages (for further discussion, see Grainger, Rey, & Dufau, 2008).

After letters and letter strings are activated or “fired”, the information goes one step further to activate phonological and even semantic information. The information passes through two routes for meaning extraction. One is through the

Grapheme-to-Phoneme route, in which phonological coding is first activated and then the semantic information is retrieved, while the other is through the lexical route, in which orthographic information directly leads to semantic meaning. The latter is especially needed for words with irregular pronunciation (Coltheart, 2000; and see Jobard, Crivello, & Tzourio-Mazoyer, 2003 for relevant information). Researchers in reading found that these two routes are probably not contradictory, but rather co-existent, and they compete as morphological information is fed to both routes and whichever processes faster wins the "meaning retrieval game".

Phonological information is important for word identification and even for comprehension during reading (Frost, 1998). Inner speech can be a persuasive support for this point of view. It's been found in many electromyogram (EMG) studies (the tool which measures the electrical activity of muscles) that less-skilled readers tend to rely on subvocalization more than skilled readers do, and the phenomenon of subvocalization becomes more prevalent as text difficulty increases. In addition, comprehension suffers when difficult texts are encountered and subvocalization eliminated (Rayner et al., 2012).

The phenomenon of inner speech does seem to make some sense, since children start learning a language through speaking and listening, while reading comes later as a relatively unnatural activity. In order to facilitate learning how to read, children are inclined to resort to sounding out words (Rayner, Foorman, Perfetti, Pesetksy, & Seidenberg, 2001, 2002). Aside from this fact, phonological information is preserved better in short-term memory (Rayner et al., 2012), which is absolutely needed for information processing and meaning integration.

Generally, as the activation of lexical representation passes a certain "threshold", a word is recognized, and the meaning and syntactic properties become available for further processing (de Groot, 2013). However, the intended meaning of a word does not naturally "emerge" every time. For words with multiple meanings, a selection mechanism has to be in place to ascertain the suitable meaning that is plausible and consistent with the whole context or the previously constructed scenario.

Rayner, Cook, Juhasz, and Frazier (2006) clearly demonstrated through fixation durations the effect of meaning selection and the cognitive effort induced. In their experiment, a word in each sentence was selected as the ambiguous stimulus. In one condition, the dominant meaning of the ambiguous word would be suitable for interpreting the sentence; in the other, the subordinate meaning would be a better fit. In addition, sentences were classified into two categories: one with a modifier to prime the meaning, and one without. The researchers found that when there was a modifier that prompted the subordinate meaning, fixation time became longer than in the neutral context and in conditions that went with the dominant meaning. Interestingly, the ambiguous word with the dominant meaning was fixated longer when not preceded by a dominant-meaning prompting modifier. The above findings suggest that a context, even just one modifier preceding the target, can affect the meaning retrieval process, not to mention the influence of the global context, which clearly sways the amount of time spent on resolving meaning issues (see Duffy, Morris, & Rayner, 1988; Kambe, Rayner, & Duffy, 2001).

In Rayner, Warren, Juhasz, and Liversedge's (2004) experiment, sentences that were 1) normal, 2) possible but somewhat implausible, and 3) anomalous were employed to examine their effects on eye movements. The results showed that, starting from gaze duration, eye movement indices were obviously affected in the anomalous condition, meaning the comprehension process was seriously disrupted, while in the implausible condition, go-past time was the earliest index to "take a hit" but did not reach statistical significance. In Rayner et al.'s (2004) experiment, the syntactic structure was fixed, confining the possible thematic roles or features of the ensuing constituent, therefore limiting the discussion to plausibility of the matching between the verb and the noun that followed.

In addition to semantic information, syntactic properties also impact the comprehension process in their own way. Frazier and Rayner (1982) successfully proved the influence of syntactic parsing. In their experiment, two sentences as stated below were shown to participants:

The girl knew the answer by heart. (Direct object)

The girl knew the answer was wrong. (Sentence complement)

Participants were found to have more difficulty reading the second sentence. Eye fixations were longer, and the number of regressions increased, suggesting that readers originally adopted a certain way of assigning roles to each constituent, which was suitable for the first sentence, but not for the other. When they found the grammatical structure of the following part did not correspond to their assumption, they had to reanalyze the sentence.

Garden-path sentences are another kind of stimulus most favored by reading scholars. McDonald and Carpenter (1981) successfully drew out the influence of idiom interpretation on participants' comprehension. In their experiment, sight translators in the study returned to former regions (in the form of regressions) when encountering disambiguating information that exhibited inconsistency with their earlier interpretation of the idiom further down the sentences. Though this experiment is more related to semantic processing, when combined as a whole with aforementioned studies, it shows that sentence processing is incremental in its nature (see Frazier & Rayner, 1982; Rayner et al., 2006).

2.2.2 Bilingual reading

The task of reading and comprehension starts from word identification to semantic selection and integration, syntactic processing, and finally to continuous mapping of current information to the mental representation and incessant verification of the consistency across perceived messages. Many factors have a role to play on the path towards comprehension in the process, including orthographic, phonological, and morphological information, semantic richness, frequency, familiarity, plausibility, syntactic properties, and even context. The same holds true for bilingual reading.

Complex as reading already is, we can fairly claim that bilingual reading is even more complicated because two language systems may be activated at the same time, though there is only monolingual input. This non-selective view of language activation has already been proven by numerous studies. For example, Dutch-English bilinguals were invited to participate in the experiments of Dijkstra, van Jaarsveld, and ten Brinke's (1998). For the English lexical decision task, Dutch-English cognates, homographs, and control words that do not share such similarities were used as stimuli. Participants reacted to cognates significantly faster than to control words. In the meantime, homographs did not bring this kind of processing benefits. This points to the facilitative effect of knowledge in a language while encountering stimuli in another language.

Dijkstra, Grainger and van Heuven (1999) conducted similar experiments to further identify how cognates, homographs, and homophones influence reading behavior. Dutch-English bilinguals were recruited, and this time the authors successfully found the impacts of each element, that is, the effects of overlap in meaning, orthography, and phonology between two languages. Significant facilitation in response time was observed for cognates and homographs, but the effects brought by homophones were the opposite, that is, participants spent significantly longer time identifying homophones than controls. According to the authors, the lack of homograph effects in Dijkstra et al.'s (1998) paper lies in the selection of stimuli. In other words, the positive homograph effects had been confounded by the inhibitory phonological influence.

In addition, Beauvillain and Grainger (1987) also found orthographical effects by using homographs in French and English. English-French bilinguals were shown a French prime word followed by an English target word (or non-word) and had to decide whether the target was a real word. The results show that non-target language seems to have been activated initially, independent of the language context. What's more, the frequency of each word in either language was more capable of determining whether meanings would be activated or not than the language mode per se, lending more clear support for the non-selective view of language activation.

Duyck, van Assche, Drieghe, and Hartsuiker (2007) used both a lexical decision task and a sentence-reading task to verify the effects of cognates on reading. The Dutch-English bilinguals showed a facilitation bonus on trials with cognates, in comparison with control trials, and response was significantly faster. Furthermore, even during sentence reading, cognates (both identical and non-identical) still sped up processing, shortening the response time at a statistically significance level. Though the third experiment of theirs only successfully replicated facilitation effect for identical cognates, it still attests to the idea of cross-language activation.

Also using sentence reading as the task is Schwartz and Kroll (2006). Spanish-English bilinguals with a high- and intermediate- L2 proficiency partook in their experiments to read L2 sentences. For the highly proficient L2 group, significant cognate facilitation was observed only in low-constraint sentences, meaning there were indeed cognate effects, but the context provided by the sentence was strong enough to suppress non-selective activation (as evidenced in high-constraint sentences). The group with less L2 proficiency generally showed similar trends. However, one thing worthy of note is that homographs did negatively influence this second group's performance. Homographs are identical in form but different in meaning between two languages, so this group of participants experienced competition between meanings during the experiment, leading to higher error rates of judgment even for highly contextual sentences.

Van Hell and de Groot (2008) looked into the influence of contextual constraint of a sentence on reading as well. Cognate and concreteness effects were examined with high-constraint and low-constraint sentences. The authors invited Dutch-English bilinguals as subjects, and they were to engage in a lexical decision task or a translation task. Participants had to first read a sentence attentively (serving as the context), and then to 1) decide whether the target word that replaced the context was a real English word or 2) translate the target word as soon as the word appeared. On the whole, cognate effects were vibrant during low-constraint sentences and largely reduced for highly contextual ones for both lexical decision and translation tasks.

Following the same line of research, Libben and Titone (2009) asked French-English bilinguals to read English sentences embedded with inter-lingual homographs and cognates in either a high- or low-semantic-constraint context. Results again mirrored previous studies. Homographs were read significantly slower than control words under both low and high sentential constraints. Cognates, on the contrary, exhibited facilitative effects, largely shortening the reading time in both types of sentences. Specifically, this study found that non-selective access of languages is sustained, even during early stages of reading for high-constraint sentences: The effects disappeared only in late-stage measures (i.e. 350-600 ms), while in the low-constraint context the impacts remained effective. Another interesting finding is that cognate facilitation, by and large, decreased as L2 proficiency increased, which may provide a hint towards understanding the relationship between linguistic competence and cognitive mechanisms.

In addition to orthography and meaning, phonological information in either language seems to be available for access (and even mandatory) during bilingual reading. Van Leerdam, Bosman, and de Groot (2009) asked Dutch-English bilinguals to read one English word and listen to a speech segment at a time and judge whether the two rhyme. Stimuli were all English words, with some having friendly neighbors (similar pronunciation) and enemies (different pronunciation) in English or in Dutch, some being distinct in its own way. The authors found that participants were often misled by enemy neighbors, and error rates for deciding whether the written stimulus rhymed with the speech segment rose significantly, supporting the view of cross-language phonological activation. This mechanism seems to transcend the boundary between different alphabets because the same phenomenon has been already documented in another study, this time regarding Hebrew-English and English-Hebrew bilinguals, by Gollan, Forster, and Frost (1997). Though asymmetric activation was observed during the experiments, the results still showed that phonology played a part in assisting participants in reading. When reading an L2 word with an L1 prime, access to cognates (words overlapping in meaning and phonology) was facilitated more than to non-cognates (words only overlapping in meaning), and the difference reached significance. On the other hand, the priming effects were not found with L2 primes. Interpretation and implications of this study should be treated with caution because it used loan words (cognates that are actually borrowed from English). According to the authors, this is inevitable because Hebrew

> was revived as a spoken language early in the 20th century after having been used solely for religious purposes for hundreds of years. For this reason, the language lacked many lexical items for modern terminology. Although some modern words were created by adapting a Hebrew root morphologically, many modern words were simply borrowed from English (Gollan et al., 1997, p. 1123)

While loan words in Hebrew are morphologically changed, the similarity between a loan word and an English foreign word is so overwhelming that, for many people, the two words are actually identical, leading to a possible L2-L2 combination in this priming task. Luckily, a similar trend in English-Hebrew bilinguals may serve to clear this doubt.

We have seen plenty of evidence that supports the non-selective view of language access for bilinguals. However, the aforementioned findings are confined to word-level phenomena. Does non-selectivity still hold when it comes to sentence reading and comprehension? This question is important and sound, since the topic of this current dissertation deals with sigh translation, in which sentence-level processing is as important as word recognition, if not more crucial.

Indeed, studies on bilingual sentence reading have reported findings in support of non-selectivity. What's more, language proficiency, age of acquisition and language exposure (environment) may all have a part to play when it comes to bilingual reading (for more detail, including findings on semantic and syntactic processing in bilingual reading, see Fernández, 2002; Frenck-Mestre, 2005; Hernández, Fernández, & Aznar-Besé, 2009).

Frenck-Mestre (1997) reported findings of several experiments, among which one instructed highly proficient French-English bilinguals to read syntactically ambiguous English sentences. The reading pattern for skilled non-native readers resembled that of native English readers. Nonetheless, the author still observed some L1 influence during the process. That is, participants hesitated when they looked at the subordinate verb in the sentences that violated lexical constraints in their native language. In another experiment, a group of beginning English-French bilinguals and a group of French native speakers both read French sentences, each with a relative clause. While the French native speakers showed preference for N_1 attachment, linking the first noun phrase appearing in the sentence with the relative clause, English-French bilinguals did not have any reliable preference but a slight tendency towards N_2 (the second noun phrase in the experimental sentences) attachment, which is typical in English. Though it might be possible that people tend to attach new information to the most recently processed element when faced with a not-so-familiar language, with another experiment reported in the same paper, the author was able to ascertain that the aforementioned phenomenon indeed came from L1 influence. In this last experiment, beginning Spanish-French bilinguals were recruited to read French sentences with a relative clause. Since the Spanish participants were less skilled in French, a logical consequence following the recency attachment strategy would be that the relative clause was linked to N_2 consistently. Yet, both Spanish-French bilinguals and native French participants manifested an identical inclination, which is natural in both the Spanish and French language — they all had reliable preference for N_1 attachment. From Frenck-Mestre (1997), we can clearly see the impact of a bilingual's mother tongue and how second language competence affects one's reading pattern (see also Frenck-Mestre, 2005 for similar findings with ERPs).

Frenck-Mestre (2002) also attested to the fact that proficient second-language readers exhibit highly resembling pattern of processing with that of native readers. According to the author, though total viewing time may be longer for L2 readers,

> proficient second-language readers demonstrate an immediate sensitivity to the same factors that influence native readers' initial progression through the sentence. Both structural ambiguity and lexical constraints influence the first pass reading times of proficient non-native readers, just as they do those of native readers (Frenck-Mestre, 2002, p. 219).

A somewhat different voice comes from Dussias (2003). Dussias invited proficient English-Spanish and Spanish-English bilinguals to fill out a questionnaire regarding choices when reading temporarily ambiguous sentences with a complex noun phrase followed by a relative clause, and to also perform a reading task to gather real-time data. As a result, both groups preferred low attachment, similar to English monolinguals but not native Spanish speakers. The results partly contradict the notion of L1 influence. One explanation provided by the author was that the cognitive demands exerted by the stimuli might have induced the minimal effort strategies (recency attachment effect). However, an alternative and likely explanation is that language exposure plays a role, since both groups of participants live in an English environment.

The role of language exposure in reading has been further substantiated by Dussias and Sagarra (2007). Two groups of skilled Spanish-English bilinguals, living in the Spanish and the English environment respectively, and an additional group of Spanish monolinguals all read structurally ambiguous sentences as those in Dussias's (2003) study. The results showed that monolinguals and bilinguals living in the L1 environment exhibited L1 processing preference for relative clauses; on the other hand, the group of people living in the L2 environment clearly showed L2 processing preference.

In addition to language ability and exposure, how age of acquisition affects language processing can be observed in Izura and Ellis's (2004) translation judgment task, in trials of which participants had to decide whether the meanings of a pair of words were identical or not. Spanish-English bilinguals participated in the experiments, and the authors found that, when translation was involved, early-acquired L2 words were processed faster than late-acquired L2 counterparts, while one's native language also determined the relative speed of meaning extraction during translation.

Up to this, point, we have seen the intricate dynamic between two languages in bilingual reading. On the word level, phonological (homophones), orthographical (homographs) information and a shared origin (cognates) may affect word identification in reading or even translation in their own way, positively or negatively. As for factors outside the boundary of words, contextual constraints, the reader's first language, proficiency of each language, the time when the reader learn certain words, or even the environment (language exposure in daily life) may influence the efficiency of reading, all due to the non-selective feature of all the language subsystems in bilingual reading.

2.2.3 Reading in sight translation

Placing the reading task for sight translation under the microscope then, it may be reasonable to speculate that all factors affecting reading for comprehension will hold true in the circumstance of reading for sight translation, since the latter is still a kind of reading. Furthermore, we can fairly expect the full force of factors that impact bilingual reading, firstly because all interpreters performing sight translation are necessarily versed in two languages to acquire information written in one language and orally translate it in another. Secondly, the fact that reformulation is involved in sight translation means the activation of both languages in rapid succession is a must, but non-selectivity might still need be avoided for efficient reading. This conscious control will inevitably consume the interpreter's limited cognitive resources.

Although we might not see cognate or homograph effects in the language combination of Chinese and English (the symbols used being highly dissimilar), as exemplified by de Groot (2013), other agents, such as the fundamental differences of syntactic rules, may still assert their presence. On top of the possible influence of bilingual reading, what may further complicate the situation is that reformulation and production in a different language is required for sight translation. This means activated linguistic information in either language might have to be suppressed quickly in one stage and deliberately solicited in another stage if interpreters are to smoothly produce oral renditions.

Will the intricate mechanism required by sight translation alter the behavior of input reception (reading), differentiating reading for sight translation from reading for comprehension? In fact, previous studies have reported rather mixed assertions. In a sight translation task (referred to as simultaneous translation in the paper), McDonald and Carpenter (1981) designed garden-path sentences, each embedded in an 85-word context, to see how two translation experts and two amateur German-English bilinguals read and parse the stimuli. The authors held that reading for sight translation resembles reading for comprehension because eye movement measures collected from their study showed that participants first read silently to understand the sentences and then began reformulation during the second pass of reading.

Shreve, Schäffner, Danks, and Griffin (1993) reached a similar conclusion with their experiment. Three groups of participants were involved. The first group, consisting of ten graduate students majoring in translation, was assigned to reading for translation, while the second group of ten M.A. or doctoral English majors was assigned to reading for paraphrasing, and the third group of 13 M.A. or doctoral psychology majors was responsible for reading for comprehension. Though all the groups didn't really do anything other than reading, the researchers believed that by simply informing the participants of the purpose for reading, reading time would be affected. The results showed that the reading purpose did exert some influence on the reading time, which was the longest for reading for translation, followed by reading for comprehension and for paraphrasing, respectively. In general, translators did read differently to some extent and there was a wide within-group variability, but this kind of reading still "has much in common with other kinds of reading, especially with that of paraphrasers, where language conversion is required by the tasks" (Shreve, Schäffner, Danks & Griffin, 1993, p. 35). According to the authors, though it could be said that translators read more thoroughly and deliberately to some extent, the reading strategy still resembled a more general reading for comprehension in much part.

On the other hand, two experiments in Macizo and Bajo's (2004) study showed contradictory findings. In each experiment, eight professional translators who did not participate in the other experiment were recruited to either read to comprehend or read to translate 96 sentences. Generally, the reading time of reading for translation was significantly longer than that of reading for repetition, indicating some other processes going on at the same time while one was "reading" the sentences (see also Macizo & Bajo, 2006; ambiguity and cognate facilitation evidently affected reading for translation, but not reading for repetition).

Göpferich, Jakobsen, and Mees (2008) sided with this purpose-driven view of reading behavior as well. Six professional practitioners and six translation students participated in the research, and four types of tasks were used: reading for comprehension; reading in preparation for translating; reading while speaking a translation (sight translation); and reading while typing a written translation (Göpferich et al., 2008, p. 106). While the difference between the mean fixation duration of different tasks was only close to significance, we still see an obvious numerical gap between any two modes. Meanwhile, the number of fixations was considerably and reliably different between all four tasks. The findings are a further testimony to the fact that various purposes for reading lead to distinct behavior.

With a similar goal in mind, Shreve, Lacruz and Angelone (2010) used two Spanish texts for participants to perform sight translation. Eye movement information was recorded with an eye tracker for comparison with the data of a bilingual reading task. In general, there were more and longer fixations during sight translation than bilingual reading. On top of that, regressions were significantly more, and larger proportion of fixations was accounted for by regressions in sight translation. Taking the above results into account, reading for comprehension does seem to differ from reading for sight translation.

Yet, Shreve et al. (2010) warned that the results have to be interpreted with caution, since the baseline study (the bilingual reading task) was silent reading. As Levy-Schoen (1981) has pointed out that reading aloud is generally a more effortful activity than silent reading and causes longer eye fixation durations, some of the significant difference between sight translation and bilingual reading might emerge as a function of reading out loud the sight-translated text, rather than of the demand imposed by the component of reformulation.

The above findings, though somewhat divergent, on reading patterns in different tasks open a window into the cognitive world of sight translation. Studying the behavior of reading in this mode of interpreting is a direct, scientific method towards understanding whether the characteristics of the mode alter the processing mechanism used in average daily reading.

2.2.4 A perspective provided by reading research: Eye-movement indices

To really understand what happens or what causes problems in the process in order to get a better picture of sight translation, we need a tool with high temporal resolution to reflect on-line changes of cognitive operations and in the meantime allowing oral rendition during the experiment. Sight translation relies on reading to gain information, and therefore an eye tracker seems to be the best fit for the time being, since some other prominent research methods are not compatible with the requirements. For example, speaking can easily confound the data collected with ERPs, while PET is rather expensive and low in temporal resolution, though accurate in pinpointing locations (see Mitchell, 2004 for further detail).

The advantages provided by an eye tracker are numerous. First of all, it directly monitors the eye. Based on Just and Carpenter's (1980) immediacy assumption and eye-mind hypothesis, when the eye fixates on a word, processing automatically begins, and the aforementioned morphological, phonological, lexical information and even word meanings, grammatical features, contextual constraints will start to come into play until all processing required for comprehension is completed. In addition, the mind (attention) will direct the movement of the eye, and hence we can be quite assured that what the reader is reading is practically what he or she is (at least partly) dealing with. Although Inhoff and Radach (1998) warned the possibilities of pre-processing of the word ensuing the fixated target word, and Radach and Kennedy (2004) also illustrated the workings of the spillover effect (the cognitive burden of the previous region carried over to the current target word), we can still be confident that monitoring the time spent on the fixated area offers useful information about the cognitive workings of the brain. For one, it's hard to imagine any reader would choose to fixate on certain places with deliberate intention to suppress all related processing. For another, even with the pre-processing and spillover effect, a large part of the time spent still reflects the processing of the current target, which stays in the foveal vision that fits best for information extraction (see also Rayner, 2009).

With an eye tracker, we are able to document moment-to-moment changes of the eyes and see on-line variations of cognitive load/effort or processing strategies, and even identify difficulties with data about *when* and *where* eyes move. Where interpreters' eyes are fixating, for how long, and how they move while producing output simultaneously are interesting enough, but equally fascinating is what they try to do when they are not speaking. With more detailed information on the whole process, we may be one big step closer to the nature of sight translation.

To understand reading, scholars in this field have designed different indices to map behaviors of the eye to different stages (first-pass vs. non-first-pass) and tried to decipher the meaning behind each phenomenon. Among all kinds of indices, duration-related indicators are of major concern, while others, such as direction of movements and word skipping, serve as supplementary information, providing a more complete picture when taken into consideration all together. Below we introduce the main indices used in reading and sight translation research.

First-pass fixation measures mainly include first fixation duration and gaze duration. These are the fixations made after an eye first enters a specified region of interest and before leaving it, and, whether alone or aggregated together, are considered as relatively immediate/early indices of cognitive processing when larger chunks are used as the unit for analyses (see for example Hyönä, Lorch & Rinck, 2003). Put separately, first fixation duration indicates "the duration of the first fixation on a target word during first pass sentence reading, irrespective of whether the target word receives one or several fixations" (Inhoff & Radach, 1998). When there is only one first-pass fixation on a target word, first fixation duration equates to gaze duration.

As the first contact with new information, first fixation duration represents adequately the pre-lexical or lexical processing of a certain word. Some scholars doubt if first fixation duration really reflects the sole influence of the currently fixated word. Their suspicion does have some merit, since we have seen empirical evidence for preview and spillover effects. Taking it into consideration, a word may possibly be processed before even being fixated, and, moreover, we may see lagging effects from earlier text in disguise of first fixation duration of the current word (Rayner, Juhasz, & Pollatsek, 2005; Tsai, Yen, & Wang, 2005).

Gaze duration is defined as “the sum of all fixations on the word prior to moving to another word” (Rayner & Liversedge, 2004). Similar to first fixation duration, gaze duration is sensitive to some pre-lexical, lexical, and post-lexical features of words and texts. Looking from a broader perspective, gaze duration still manifests relatively early stages of processing; but when taking words as AOIs, one may find some factors taking its toll on gaze duration but not on first fixation duration (“later” effects) and vice versa (“early” effects that appear immediately but do not last long). Gaze duration shows a linearity pattern of variation as a function of word length, since some re-fixations are required for longer words compared to short words. At this point, it’s worth mentioning one closely related index: re-fixation rate. Re-fixation rate means “the ratio of the frequency of multiple target fixations over the frequency of multiple and single target fixations during first pass reading” (Inhoff & Radach, 1998; see also Tsai et al., 2005). Re-fixation rate can easily manifest impacts of word frequency, word difficulty, word length, and initial landing location.

In translation and interpreting studies, gaze duration may be a great index reflecting the efforts needed to process words (McDonald & Carpenter, 1981) or even to start reformulation (cf. Chen, 2013). However, the question is, the indicator alone cannot tell us what the interpreter is really doing. For example, in Chen’s (2013) experiments, early fixation time of experts was found to roughly equate to that of novices, contrary to general assumption that experts process information faster. Nonetheless, the quality was found to significantly supersede that of novices as a group. Chen inferred that the experts might have performed a greater deal of work within similar amount of time.

Mean fixation duration is probably one last important first-pass index. It is seen as the mean duration of all fixations that appear during first-pass reading. As informative as other figures, mean fixation duration is susceptible to distortion. For example, fixations on a more difficult word may be more than those on an easy word, but with shorter duration for each fixation. In this way, the mean of multiple fixations has a chance to be less than single fixation duration on an easy word.

Early indices are affected by orthographic, phonological, morphological properties, word frequencies, personal familiarity, and even contextual constraints, to name just a few (further detail see Rayner, 2009; Rayner et al., 2005; Tsai, Lee, Tzeng, Hung, & Yen, 2004). It's wide consensus that first-pass measures mostly reflect the cognitive engineering of word identification and lexical processing.

All the measures that emerge after the eye first leaves a target region will be labeled as second-pass reading. A few fundamental indices used for later stages of processing include rereading time, go-past time, and total viewing time, plus an important and related probability index: regression rate. Rereading time, according to Hyönä et al. (2003), is computed by adding all the subsequent fixations other than first-pass reading on a target word, representing later-stage processing after the word is identified.

Go-past time (in some studies also named regression path duration) is defined as "the time from first entering a region until moving past that region forward in the text" (Rayner & Liversedge, 2004). Some scholars use this index to capture the effects of some phenomena which are generally not easy to be identified through standard measures, including syntactic ambiguity, inconsistency, or the amount of processing load accumulated from earlier to the current region (see also Vonk & Cozijn, 2003). Go-past time may be especially useful in studies of sight translation, since some structural disparity and the different habit of placing main or subordinate message in the front between two languages may force the interpreter to go back when s/he has to restructure the whole sentence or digest again the heavier-than-expected information that need to be knitted closely with the current content.

Regression rate, which stands for the percentage of readers moving their eyes back to previous regions, is tightly bound with go-past time and conveys similar messages. An issue with go-past time is that readers don't necessarily go back for clues to solve problems at hand. They sometimes go further down, trying to gather more information to see if they can construct a more coherent structure of the content (see also Rayner & Liversedge, 2011).

Lastly, total viewing duration (or total viewing time) sums up "all fixations in a given region" (Rayner & Liversedge, 2004). Clearly as the name suggests, total viewing time represents the assumed total cognitive effort expended on a target word or region. In reading, larger amount of total viewing time either indicates more difficulty encountered or more important information embedded. Since sight translation involves reading as a necessary part of the task, the same index has the same potential of identifying the culprit that prevents interpreters from completing their task with flying colors. Further, when combined with other indicators, we might even find whether interpreters and readers, with different goals in mind, mostly care about or are boggled down by similar constituents or not.

In order to outline the cognitive efforts in different stages during sight translation, two first-pass indices (first fixation duration and gaze duration) and three non-first-pass indices (go-past time, rereading time, and total viewing time) were selected for this study. In this way, we may be able to see word retrieval in early stages, meaning integration and information searching (or problem solving) in later stages, and the total time spent on words that reflect efforts to achieve appropriate mental representation.

2.3 Linguistic differences between Chinese and English

Chinese and English belong to two distinct language systems. To begin with, English is an inflectional language with words composed of individual letters and a space serves as the boundary between every two neighboring words, while Chinese is non-inflectional with each Chinese character constituting words of different length without a clear boundary in between (Frost, 2012; Huang, 2009). Since affixations indicating tense, active or passive status, and other functions are not available in Chinese, word order seems to be of critical importance for users of this language (Jiang, 2009). As is claimed by Huang (2009), "Chinese is characterized by non-inflection and must express various syntactic and semantic relations through the use of function words and functional manipulation of word order" (p. 3).

In addition to this fundamental difference, whether modifiers should precede or follow the modified is considered comparatively free in English but strict in Chinese. Other issues regarding the two languages have also been frequently discussed by scholars in linguistics, including the subject- vs. topic-prominence or the sequence of thematic organization (for example, see Chu, 1998; Hoosain, 1991; Li & Thompson, 2009).

In order to systematically reflect the similarities and differences between languages, scholars often resort to four major typologies, which are principal branching direction (PBD), head parameter, topic-prominence vs. subject-prominence, and morphological typology (Chen, 2006a). Among the aforementioned four parameters, principal branching direction is the focus of this study, since in sight translation the principal branching direction is the most obtrusive feature that jumps out at the reader quite often when one reads written text in a non-native and structurally distinctive language.

The concept of principal branching direction refers to the linguistic structures (modifiers) that are unmarked for most native speakers, and the concept basically covers the relationship between the modifier and the modified (also called "head"), that is, the relative position of the two components in one or multiple clauses. Here the "structures" mentioned are for the most part adjectives and adverbs on the word and phrase level, and also relative clauses and adverbial subordinate clauses (see

Chen, 2006a; Diessel, 2004). In terms of adverbial subordinate clauses, most common types include temporal, conditional, concessive, and causal clauses.

This dissertation confines itself to the use of the "modifier" and the "modified" because there is another criterion for contrastive linguistic analysis (i.e. head parameter) that also uses the word "head" but focuses on the relative position of the head and the complement (Chen, 2011).

There are two categories under the typology of principal branching direction: left-branching (the modifier preceding the modified) and right-branching (the modifier following the modified). While Chinese is mostly left-branching (Huang, 2009; Yang, Perfetti, & Liu, 2010), English is centered around the right-branching principle on the phrase and sentence level and is only left-branching on the word level (Chen, 2006a). Consider the following examples:

1. Sentence level:

(1) He is the attorney {whom I met yesterday}. (Relative clause)

<Modified> + <Modifier>

他就是{我昨天遇到的}那位律師。

<Modifier> + <Modified>

(2) He went home early {because he was sick}. (Adverbial subordinate clause)

<Main Clause> + <Subordinate Clause>

{他因為生病了}，所以提早回家。

<Subordinate Clause> + <Main Clause>

2. Phrase level:

(1) The girl {across the street} is pretty.

<Modified> + <Modifier>

{對街的}那個女生很漂亮。

<Modifier> + <Modified>

(2) I like to study {in the library}.

<Modified> + <Modifier>

我喜歡{在圖書館}讀書。

<Modifier> + <Modified>

3. Word level:

(1) He is a {smart} kid.

<Modifier> + <Modified>

他是個{聰明的}孩子。

<Modifier> + <Modified>

(2) {Stinky} tofu is my favorite.

<Modifier> + <Modified>

{臭}豆腐是我的最愛。

<Modifier> + <Modified>

The examples in the previous page clearly show that Chinese and English are quite similar on the word level. However, these two languages start to differ as the modifier grows from a single word to a phrase or a clause, with Chinese becoming stricter when considering the relative position of the modifier (for relevant discussion, see O'Grady, 2011).

As a matter of fact, the stringent left-branching feature of Chinese is closely related to its "topic-comment" structure (relevant discussion see Young, 1982; Kirkpatrick, 1993). As Kirkpatrick (1993) put it, Chinese speakers prefer to offer supporting material or background information (topic) before delivering their main points (comment). Empirical studies in Chinese have also corroborated the notion of Chinese as a left-branching language and the relationship between this notion and the topic-comment structure. Wang (2006) collected a large sum of spoken and written data in Chinese from naturally occurring conversations, broadcast interviews and call-ins, and magazine articles to analyze whether the adverbial clauses (temporal, conditional, concessive, and causal in the study) precede or follow the clause they modify. The results demonstrated that an overwhelming number of adverbial clauses appeared initially, even in spoken discourse. The only exception might be the causal clauses. The percentage for this type of clause to occupy the initial and the final position respectively was roughly similar.

Biq's (1995) similar findings seemed to further prove this right-branching phenomenon for causal clauses, though contradicting what Young (1982) and Kirkpatrick (1993) found. However, the data used by Wang and Biq were relatively informal spoken data, which diverged from the more formal written data examined by Young and Kirkpatrick. This partly explains why the left-branching principle in Chinese was violated. In addition, Wang also observed that the final causal clauses might have played a different role from those occupying the initial position, admitting that "final causal clauses that occur after ending intonation in conversation are more often coordinate clauses rather than subordinate ones, which comment on a cause, relevant to the preceding clause" (Wang, 2006, p. 79).

On the other hand, English seems to allow some flexibility when choosing the appropriate location for adverbial clausal modifiers, though functions differ when occurring at the initial and final position respectively (Berg, 2009; Ford, 1993; Ramsay, 1987). Prideaux (1989) analyzed 1331 pages of texts of dissimilar styles, from "oral-like" to a formal tone, to see how each temporal adverbial clause (starting with *after/before*) combines with the main clause it tends to modify. The results denoted that the number of initial subordinate clauses appeared to be less than that of final subordinate clauses. In addition, the author detected a different function provided by initial adverbial clauses, which are marked structures in English, and claimed that "a marked structure would seem to be a natural indiction*[sic]* of a discourse break, providing a guidepost to signal the completion of one discourse package and the advent of a new direction, theme or topic" (p. 38). Prideaux and Hogan (1993) asked participants to provide oral and written descriptions of a short film clip to see how adverbial subordinate clauses (e.g. when, since, before, after, while, as) were used. Their findings replicated Prideaux's (1989) and also found that the adverbial clauses were placed before main clauses to bring listener/reader's attention to a new discourse unit.

Still, it is widely acknowledged that English is mostly right-branching in its nature (e.g. Clark & Clark, 1977; van Riemsdijk & Williams, 1986, p. 211). As Prideaux and Hogan (1993) put it, "in English the order Main Clause (MC) followed by Subordinate Clause (SC) is unmarked, with the order SC+MC being the marked member" (p. 398). Specifically, relative clauses and phrasal modifiers tend to appear on the right side of the modified word as a natural sequence (see also Odlin, 1989).

The contrasting feature of principal branching direction between languages may cause problems for language acquisition (Chen, 2006a), as previous studies have sufficiently proven. For example, Schachter (1974) examined the acquisition of English relative clauses by speakers of Arabic, Persian, Japanese, and Chinese. The author contended that, though errors in producing relative clauses were much fewer for Chinese and Japanese speakers, the reason behind the phenomenon probably did not come from the fact that they really mastered the English language. On the opposite, users of these two languages tried their best to avoid using relative clauses, of which the direction of branching is in direct contrast with their mother tongue. In addition, Chen (2006b) recruited Chinese ESL learners and Spanish ESL learners to unveil the possible influence of their native language on English and vice versa and the role played by their English proficiency. The results clearly showed the influence of L1. Native speakers of Chinese produced a significant number of subordinate-main sentences (modifier-modified structure), violating the norm of English, in which the main-subordinate structure prevails. Interestingly, native speakers of Chinese in an English environment also manifested the sign of backward transfer, that is, they tended to show the sentence pattern preferred in English when creating sentences in Chinese. Chen (2007) again investigated the influence of L1 and the role of L2 proficiency. This time, English and Korean learners of Chinese were invited to participate his experiment. The results corroborated what the author found in his previous study. Native speakers of English still showed a tendency to choose the main-subordinate structure over the opposite structure they were supposed to use when producing sentences in Chinese.

With English-Chinese sight translation in mind, we have reviewed some major differences between the two languages. Fundamental disparities aside (i.e. characters vs. letters; non-inflectional vs. inflectional), Chinese is basically a topic-prominent language, whereas English is subject-prominent. However, an even more conspicuous contrast is their preference in the principal branching direction for modifiers: Chinese adheres rigidly to the principle of left branching on the word, clause, and even sentence level, while English is more flexible, though mostly right-branching, except word-level modifiers. This directly contrasting syntactic structure may inevitably pose challenges when performing English-Chinese sight translation, since the information about antecedents have to be held onto until modifiers (some of which may be unbearably long, such as relative clauses, especially in formal, written speeches) are digested, reformulated, and rendered properly. Chinese's rigidity is also the reason the current study made Chinese the target language, so that the participants would not be able to follow the order of the source text to save resources for other cognitive operations.

2.4 Effort Models

Simple as the definition of sight translation is, in order to really grasp what cognitive activities are possibly involved, we might have to turn to Gile's (1997) Effort Models. The models look at different modes of interpreting from a cognitive perspective and try to account for "problem triggers" during interpreting and related issues, such as interpreters' processing capacity and the allocation of cognitive resources, and also the meltdowns of performance. Three modes of interpreting were in the scope of Gile's discussion: simultaneous interpreting, consecutive interpreting, and sight translation.

Simultaneous interpreting is a task in which the interpreter has to first listen to source speech and then to reformulate the message for production, while at the same time still taking in new information. According to Gile (1997), the whole task can be broken into the element of listening and analysis (L), production (P), memory (M), and coordination (C). The listening and analysis effort deals with all comprehension-related activities, and the production effort handles the planning of speech and the final execution; in the meantime, the memory effort is needed so information does not disappear after it first comes into the interpreter's attention and before it is reformulated and successfully delivered in appropriate structures allowed in the target language. Lastly, the coordination effort is required so that the interpreter can properly allocate cognitive resources and avoid a sudden breakdown of functions. As you can imagine, processing capacity available for each element should always supersede the amount of required capacity to keep the whole process functional.

The same idea applies to consecutive interpreting, but a fundamental difference lies in the fact that speaking and listening do not occur at the same time. The interpreter first listens to the speaker and takes notes, so efforts needed in the first phase include listening and analysis (L), memory (M), note taking (N), and coordination (C). After a certain period of time, the interpreter then deliver the translation while the speaker stop for a moment, so this second (reformulation) stage includes remembering (Rem), since one has to retrieve information from the memory, and also note reading (Read) and production (P).

Effort Models have a wide range of influence upon the discipline of interpreting studies and the field of interpreter training, and the tightrope hypothesis along with it has been substantiated (more or less indirectly) by many studies. The tightrope hypothesis predicts that inadequately more concentration on a certain part of the text lead to disastrous outcomes following or around the "ground zero" (see also Lee, 2006; Mizuno, 2005). However, these models are still too vague. Gile (1999) himself admitted that Effort Models were not established to speculate the information-processing flow or the mental structure of the interpreter in the process of interpreting, but were only to highlight the components that play a part as cognitive operations are executed. As Setton (2001) also pointed out, while cognitive overloads can sometimes show unexpected and largely unknown delayed destructive consequences, a more fundamental problem is that "hypothesized load factors, or 'problem triggers', are not sufficiently specified to correlate effort (were it to be measured) with discourse or environmental events. It is not clear how capacity is measured and what constitutes a load" (Setton, 2001, p. 10).

I would like to add to Setton's view that every component in Gile's models is too complicated in its own way to be regarded as equal, and those components (or the cognitive load exerted) alone cannot fully explain the final outcome of interpreting. Pym's (2009) reanalysis of Gile's (1999) raw data has further corroborated our view. According to Pym, participants in Gile's experiment actually took into consideration the factor of "risk level" during interpreting. In other words, in addition to the components listed in the Effort Models, interpreters actually kept an eye on how each choice they made would jeopardize the completeness of the original messages.

Gile (2009) later on expanded his model for sight translation to include memory (M) and coordination (C) as well. However, he did not specify the roles these two components assume, and the revision came with a proviso: the amount of effort needed is still less than that for simultaneous and consecutive interpreting unless the language pair involves two distinctive syntactic structures. Notwithstanding, the last sentence happens to serve English-Chinese sight translation well, since these two languages, as has been introduced, are in direct contrast with regard to the principal branching direction of ubiquitous modifiers in sentences.

2.5 Sight translation and its uniqueness

Sight translation is defined by Gile (1997) as a simplified mode of interpreting and a skill that simply consists of two components: reading and analysis (R) and production (P). The author acknowledged that sight translation does impose some difficulties. For instance, the ever presence of the source text may induce interference, and sight translators are generally expected to quickly read the text and reformulate the output as if they were simply giving an original speech in another language, while concurrently "reading ahead" to maintain the continuous flow of information (see also Chen, 2015; Göpferich et al., 2008). However, Gile still contended that sight translation is generally an easier task in comparison with other interpreting modes, even with memory and coordination efforts required.

Gile's assertion is still open to debate, as far as English-Chinese sight translation is concerned. First of all, it goes without saying that there will be a heavy burden on the English-Chinese interpreter to reformulate a long embedded sentence (which is quite common for sight translation tasks) while deciding an appropriate order for constituents in the sentence in another language, and at the same time producing the reformulated unit out loud. And let's not forget that the interpreter has to continue to receive new information due to the listeners' expectation for sight translation to be as natural and fluent as monolingual communication. Secondly, observable outputs aside, we have no knowledge of how sight translation is or should be conducted, or how cognitive resources should be allocated or balanced among different tasks (comprehension, memory, production, and coordination in this case), let alone judging whether sight translation is easier. This points out the necessity of this study, since we are still floundering in immense output data, dying to find out the nature of sight translation through extrapolation.

Much empirical research seems to confirm our suspicion. For example, as noted by Agrifoglio (2004), it's precisely the presence of the text that continues to impose visual interference, making the interpreter easily fall into the trap of remembering exact words or phrases instead of looking for the gist of the original text. Therefore, producing output as naturally as speaking directly in the target language is more difficult in sight translation. Chang (2008) further corroborated Agrifoglio's findings. In her experiment, most participants agreed that source language structures were a huge hindrance that prevented smooth and correct rendition, and the researcher contended "the subjects appeared to be undoubtedly influenced by the surface structure of English while doing English-Chinese ST." (Ibid. p. 75)

Furthermore, sight translation is unique in that information comes in through reading, not listening. As Chafe and Danielewicz (1987) adequately explained, written and oral texts are different in that the former is normally packed up with more complex vocabulary and sentence structures, including a mixture of compound and complex sentences, together with a tendency to have widely varied use of diction. In other words, writers have more luxury to spend time contemplating on the rhythm of the sentences or the twists for the unfolding of messages and avoiding repetition of the same sentence patterns or words, largely increasing the difficulty of written texts. On the contrary, oral speech often resorts to repetition to strengthen the influence, uses more popular expressions or idioms, and tends to shorten each speech segment so as to ensure that the meaning gets across easily and interaction with the audience can appear in a more real-time manner (for a comprehensive overview, see Akinnaso, 1982; see also Biber, 1988; Stubbs, 1983).

Under the impact of features of written discourse, together with the quality requirement, tight time constraint, and continuous interference from the source text, the fundamental differences between the linguistic structures of languages may possibly serve as "the last straw" that breaks the camel's back, as has already been proven by previous studies (e.g. Agrifoglio, 2004; Chang, 2008).

2.6 Traditional studies on sight translation: findings and limitations

Her (1997) conducted an action research to investigate binary[1] errors produced

[1] Pym (1992) referred to binary errors in translation as mistakes that clearly misrepresented the intended meanings of words, while non-binary errors were still correct interpretations on the most basic meaning level, being only less appropriate in a certain context.

by students performing sight translation and provided possible reasons for the errors and further suggestions for teaching. Seven M.A. students majoring in translation and interpreting participated, and through analysis of the output, categories of errors were categorized as follows: lack of common sense/background knowledge; compounds and phrases; miss the humor/stylistic elements; wrong assumption; syntactic difficulty; unfamiliar word/usage/terminology/ made-up words; cultural reference; carelessness; and lastly, collocation (Her, 1997, p. 111). Though descriptions of errors and examples for each category are in detail, and suggestions for teaching are indeed invaluable, frequencies of each category are not provided. Thus, we are unable to see a specific trend or pattern of errors, except that all the above categories are errors that may occur. In addition, the research design is rather vague. The texts chosen weren't even slightly controlled, and categorization of errors lacked rigorous standards. Furthermore, a large part of the data came from students' performance in class; yet, the output wasn't recorded and transcribed but simply recorded in the form of notes jotted by the researcher on the spot.

Viezzi (1989) chose eight texts, half in English and half in French and 600 words each on average, to test the retention rate of information in the mode of sight translation and simultaneous interpreting into Italian, the mother tongue of all the participants; listening and reading were also included for comparison. The retention rate was decided by ten multiple-choice questions, which were carefully designed to prevent the participants from guessing the answer based on background knowledge. It turned out that listening and reading shared similar retention rates, both for the

English-Italian and the French-Italian task; on the other hand, sight translation showed a much lower retention rate than simultaneous interpreting in the former combination, and the difference was much smaller in the latter. Viezzi (1989) contended that the disparity between the syntactic structures of two languages might have a say in how much information one can retain, on account of the fact that there was no obvious difference in the French-Italian task. Furthermore, sight translation seems to present larger interference because the retention rate was much lower.

Nonetheless, several drawbacks have undermined the validity of Viezzi's (1989) research. In addition to pure percentage without further statistical analysis, the scores of undergraduate interpreting students were mixed with those of professional interpreters within the same group. Furthermore, the texts were not meticulously selected to avoid possible inconsistency of the difficulty level between texts.

Lambert (2004) put sight translation, sight interpreting, and simultaneous interpreting together to examine the relative difficulty of the above three modes of interpreting. Fourteen college students with similar background and training were called upon to participate. The test material for each mode all came from the same 20-minute speech, and the testing procedure was the same for everyone: sight translation first, followed by sight interpreting, and simultaneous interpreting came last. Three raters rated all the participants' output independently, with an inter-rater reliability reaching .93. Sight translation on average scored highest (82.43), and sight interpreting ranked second (82), followed by simultaneous interpreting (69.5), but the only significant difference appeared between the latter two. It seemed with written reference quality was more easily maintained. Nevertheless, this assertion cannot be accepted too readily, for we know nothing of the evaluation criteria used in this study except from a single score. What's more, the uncontrolled texts, together with a fixed testing sequence, might have introduced fatigue as a confounding factor into the final result. As acknowledged by the author, participants had ten minutes to preview before performing sight interpreting, making test conditions unequal. What would further skew the result is that, if the speaker sticks tightly to the manuscript, chances are that participants could still attain a high score even when completely lagging behind and deciding to focus entirely on the text.

Agrifoglio (2004) compared performance and obstacles during simultaneous interpreting, consecutive interpreting, and sight translation by six professional freelance interpreters with at least nine years of experience. The materials for the three modes were taken from different parts of the same report. On average, meaning accuracy was the highest for sight translation (75%), as was expected due to the presence of text, but what was surprising was that the error rate for expression was also the highest (75%); in comparison, the meaning accuracy and expression failures both dropped to 37% in SI and 24% in CI. The results seemed to prove the claim that continuous presence of text interrupts fluent renditions, eliciting unnatural use of expressions in another language. Unfortunately, the percentage used was not comparable across different modes, for errors were calculated independently and meaning and expression failures added up to 100% in each mode. In this way, a cross-mode comparison would be misleading, for one expression error out of two in one mode would be equal to ten expression errors out of 20 in another mode. Besides, drawing materials from the same source might lead to learning effects (participants will become more familiar with the content as the experiment proceeds further), and the author did not specify whether this intervening variable was controlled. Lastly, evaluation standards and how errors were counted as meaning or expression errors were not indicated.

That said, Chang (2008) still managed to corroborate Agrifoglio's finding with a sight translation experiment on nine Chinese-native graduate school students majoring in translation and interpreting and a questionnaire inquiring about perceived difficulties. Of all the participants, 77.8% agreed that, when engaging in English-to-Chinese sight translation, source language structures were a huge hindrance that prevented smooth and correct rendition. In terms of Chinese-to-English sight translation, "most of them (66.7%) agreed that carelessness…relative clause, and long sentence may be considered as difficult aspects while doing ST" (ibid. p. 75). If sight translation is simple as that pointed out by Gile (1997), what can explain the prevalence of both binary and non-binary errors?

As a matter of fact, time may be a crucial factor that comes into play. Though rarely would interpreters be timed, it's almost rules of thumb that the sight translation output should sound as if interpreters were simply reading a written document out loud instead of translating for others (Angelelli, 1999). Consequently, interpreters may feel compelled to minimalize the silent period before starting to interpret, all too often letting some errors "fall through the cracks," as exemplified in Chang's research.

Yet, we should still interpret Chang's results cautiously. First of all, the questions included in the questionnaire overlapped to a large extent and were too vague for readers to identify difficulties imposed by sight translation. Secondly, quality assessment lay subjectively in the hands of the researcher with no clear standards for categorization of errors, not to mention the fact that the speculation of the reasons for making mistakes were solely based on the product of sight translation provided by students.

Some other studies aimed to understand how training and expertise exert impacts on the performance. A total of nine participants took part in Lee's (2012) research, among which three were first-year postgraduate interpreting students, three were in the second year, and the other three were professional practitioners. All participants had six minutes to preview the only research material, a 600-word excerpt of a speech, before sight-translating it. Three raters then rated the output in terms of accuracy, delivery, and target language expressions. On average, the professionals spent less time finishing the task, and with less major errors, although the number of minor errors was roughly the same as that of student interpreters. In addition, professional interpreters consistently had fewer target language expression errors. It seems that training or experience does affect sight translation performance. However, a few caveats should be kept in mind. First, the fact that there were huge differences within each group was not addressed, and statistical analysis was not available to ascertain whether there was any significant inter-group or intra-group disparity. Another is that raters seemed to differ to a large extent about the rating criteria, while the selection of participants did not adhere to a more objective standard, either.

Chiang, Kuo, and Chen (2009) provided another insight into the effects of training and directionality on the performance of sight translation. A total of 20 interpreting students were divided into two groups; nine had received training for one year, and the other 11 just began their study in the interpreting program. All participants had to perform two-way sight translation. The text was a 70-word mission statement of a school department written in English for English-to-Chinese sight translation task, and for the Chinese-to-English task, the same mission statement was first translated and then used as the test material. Participants uniformly read the English text aloud and then translated it into Chinese; afterwards, every participant went through the same procedure with the Chinese version of the same text prepared by the researchers. According to the authors, there were significantly more pauses, among which more were "inappropriate", during the C-to-E task than the other way around, though average pause duration remained roughly the same. As for the influence of training, senior students only showed significantly shorter pause duration in E-to-C sight translation. The result wasn't replicated in C-to-E task. In terms of the number of pauses, including "inappropriate" ones, senior students' were significantly fewer than those of less experienced students'. Despite the seemingly positive effect of training, the test result is indeed questionable, for the materials in both directions are identical, and the same procedure is followed by every participant, with E-to-C coming first and C-to-E task second. This may lead to learning or sometimes hindering effects that obscure the actual performance of students. On top of that, the duration and number of pauses should not be seen as an equivalent of quality. All things aside, it was unknown that how a pause was defined or what kinds of pauses were deemed inappropriate.

In general, sight translation has been proven to be a challenging interpreting mode in its own way. It's a shame that the aforementioned studies on sight translation are solely based on product analysis (and mainly error analysis), comprehension-checking questions, duration of output, subjective questionnaires, or interviews to enrich our knowledge about the nature and the process of sight translation and factors that may have impacts on the strategies or the oral output. The deficits of these research tools and methodology, which prevail in traditional sight translation studies, mainly lie in the inability to gain access to the interpreter's moment-to-moment eye behavior, and therefore are not able to reveal what really happens during the process or what causes problems.

2.7 Insights from cognitive psychology and interpreting studies

2.7.1 Psychological investigation into interpreting

Psychologists have been attracted by the topic of interpreting for quite some years for its mysterious nature and huge overlap with issues frequently encountered in cognitive psychology, including memory, bilingualism, language comprehension and production, attention, and even complex skill performance. Among various interpreting modes, simultaneous interpreting perhaps receives attention of the most, since it's an intricate task that is possibly one of the hardest cognitive tasks imaginable in that, during most of the time, the interpreter will be listening, reformulating, and speaking simultaneously. This is supported by Chernov's (1994) study, according to which listening and speaking overlapped around 70 percent of the time.

Simply from the above description, we may be well convinced that simultaneous interpreting exerts great demand on the limited cognitive resources that have to be divided for the storage of information, the operations of comprehension, reformulation, and production, and the monitoring of one's own output. The high mental load and stress carried by this task has been substantiated by findings of larger pupil dilation in this interpreting mode than in shadowing and pure listening (Hyönä, Tommolo & Alaja, 1995) and immediately raised blood pressure and an elevated heart rate amid the task (Klonowicz, 1994).

With this immense cognitive load and the necessity for synchronous processing, interpreters are forced to receive the input and quickly reformulate it as soon as a meaningful unit is heard, creating a phenomenon referred to as the ear-voice span (EVS), while still taking in more information on the fly. Generally the EVS lasts between 4 and 5.7 words (or meaningful units, see Goldman-Eisler, 1972) or 2 seconds of time lag (relevant discussion see Christoffels & de Groot, 2005). Goldman-Eisler (1972) invited six professional interpreters to document how messages were segmented to make simultaneous interpreting possible and found that, 90-95 percent of the chunks consisted of a noun phrase plus verb phrase and centered on the predicate; in addition, interpreters did not mimic entirely the speaker's segmentation of input, since 90 percent of the time reformulation was initiated before a natural pause.

What complicates the thing is that activation of both languages during simultaneous interpreting is necessary — one has to comprehend the input and monitor his or her own output (which becomes a kind of input as well) at the same time on the reception side, while consciously inhibiting the non-target language on the production side to avoid code-switching (for evidence of self-monitoring, see Isham, 2000; Lonsdale, 1997). It is hypothesized by scholars that, in order to achieve the aforementioned feat, one needs a sophisticated control mechanism in place (de Groot & Christoffels, 2006).

Green (1998) introduced the concept of schema and a Supervisory Attentional System, similar to a monitoring mechanism that constantly checks the suitability of current schema and whether one's goal for the task at hand is met by corresponding performance. In the context of language tasks involving two languages, activation or inhibition of the schema of different languages are required in order to choose the right target language for output (comprehensive discussion see Green, 1998). Grosjean (2001) broadened the view of inhibition and activation by stating that bilinguals adjust the activation level of each of their language repertoire to fit the communicative need of the setting, introducing the concept of language mode continuum. At the one end, language users would maximally activate one language and completely suppress another should the setting requires only one language to be used; at the other end, two languages would be equally activated. For simultaneous interpreting, which falls in the latter category, "attention is dynamically assigned to the different attention-demanding components of the task" (de Groot & Christoffels, 2007, p. 35). Due to the fact that human's attention is biologically limited to one thing at a time, repeated and fast switches among reception, suppression, monitoring, reformulation, and production eat away cognitive resources and lead to fatigue quickly (Levitin, 2014), and thus the task of simultaneous interpreting is especially delicate in nature, since one minor misjudgment of resource allocation could tip the balance and lead to poor performance. Darò and Fabbro (1994) bore witness to the demanding feature of this task. In their experiments, recall of text following interpreting was worse than after pure listening. What was more, memory span for digits was significantly poorer after simultaneous interpreting than after listening (with and without articulatory suppression) and shadowing.

Some questions arise naturally following preliminary studies on simultaneous interpreting: Is this skill exclusive to interpreters only? Or can we say that actually every bilingual has the same skill, so it's only a matter of quality? Furthermore, what differs between professional interpreters and average bilinguals, so that the former can accomplish this task with flying colors, while most of the latter suffer? Dillinger (1994) recruited eight bilinguals with no interpreting experience and eight interpreters with 8.5 years of experience on average, attempting to identify whether the two groups differ quantitatively or qualitatively when performing simultaneous interpreting. Results showed that experienced professional interpreters did not seem to adopt different methods from the inexperienced when it comes to deciphering messages. The study suggested that interpreting might not be a special set of skills that need to be acquired (see also Christoffels, 2004). In the study, comprehension process stayed the same as in average listening or other activities, but training and experience did contribute to a 17 percent efficiency boost in the study.

Statements of Dillinger's (1994) about bilinguals being endowed with skills for interpreting were indirectly corroborated by Christoffels, de Groot and Waldorp (2003). These authors further asserted that lexical retrieval capability and working memory capacity were both positively related and also independently contributed to interpreting performance. Specifically in this study, which included an English-to-Dutch simultaneous interpreting task, reading span in English and Dutch-to-English word translation were the two variables that had a direct effect on simultaneous interpreting performance. Picking up the same topic, Christoffels, de Groot and Kroll (2006) looked into professional interpreters' and untrained bilinguals' linguistic competence and working memory with various language and memory tasks (in both Dutch and English). The authors found that, while linguistic performance was similar, interpreters' working memory was significantly larger than that of average bilinguals. Hence, working memory may serve as a distinct attribute that makes interpreters' performance stand out during simultaneous interpreting. A thing worthy of notice in this study is that the between-group gap in L2 working memory was even larger than that in L1, suggesting highly proficient capabilities to temporarily store and process information may be essential to the successfulness of this interpreting mode.

However, Padilla, Bajo, and Macizo (2005) raised divergent opinions. With interpreters and two groups of untrained bilinguals (with high and low memory span) participating in the experiment, the authors found that interpreters did not differ from high-memory-span bilinguals in the dual-task setting, except being the only group unaffected during the auditory suppression task. Domain knowledge, which was linguistic competence here, was thus suspected as the central pivot of interpreting performance. Liu, Schallert, and Carroll (2004) also joined this line of argument. While working memory, linguistic competence, and cognitive abilities were similar for professional interpreters and two groups of interpreting students with different length of training, experts in their study were able to render input with higher accuracy and, more importantly, were more adept at sifting input for more crucial idea units. The results point to the possibility that experts may possess strategies or knowledge specific to simultaneous interpreting, rather than surpassing others in terms of general cognitive attributes.

Working memory and interpreting strategy seem to stand out as two interesting factors that attracted some attention in the discussion above. Will the same factors affect sight translation performance? At first sight, the working memory needed in simultaneous interpreting appears totally different from the one needed for sight translation, since the former uses auditory channel as the main source of information, while the latter relies solely on reading. However, reading also requires decoding and retrieval of phonological information (especially in English) for meaning extraction and integration, as we have already mentioned. As Brady (1991) has noted, better working memory allows readers to effectively and efficiently use phonological information, leading to a better reading outcome. The claim was supported by some studies reported in Savage, Lavers, and Pillay (2007) as well. However, the authors warned that research on the relationship between working memory and reading was still limited, and while it was already proven that working memory did impact general language processing, the exact influence on reading required further investigation.

Still, it's not entirely ridiculous to hypothesize that the factors (working memory and strategy) affecting simultaneous interpreting may have similar impacts on sight translation, since these two modes differ almost only in the channel for input reception. Yet, so far only sparse studies have followed this path of investigation into the topic of sight translation, hence a justification of adding both elements into the current study.

2.7.2 Focused effort on sight translation

As noted by former studies such as Agrifoglio (2004) and Chang (2008), unnatural expressions in the target language and terrible quality of outputs are not necessarily unusual, if not highly frequent, when performing sight translation. However, for a really long time, most of the relevant studies have been based on product analysis or quality rating. As a consequence, we know "what" the end results are, but little is revealed about "how" sight translation is done, "where" interpreters encounter problems, or even how they "read" in the process. Do interpreters read texts the way similar to reading for comprehension? Or do they somehow develop a unique approach specific to sight translation? How do they juggle between reading and reformulation? And how does training and experience affect their eye behavior?

In order to answer these questions, scholars start to apply research tools and methodology typically used in other fields (mainly cognitive psychology) to study sight translation. For example, two translation experts and two amateur German-English bilinguals were recruited to participate in McDonald and Carpenter's (1981) study in order to find out how idioms are parsed during sight translation. A total of 12 idioms were selected and each put in a context of 85 words. The context consisted of a priming cue suggesting whether the idiom should be interpreted in the literal or idiomatic sense, and the priming cue might or might not be consistent with a disambiguating sentence following the idiom, which was the Area of Interest (AOI) in this study. The paragraphs with the literal meaning of the AOI intended were used as the baseline, and the data of fixations and eye movement patterns were collected to indicate how people parsed sentences into meaningful units and sight-translated them. All in all, participants first read the text until a meaningful unit was found, followed by a regression to reread and orally translate the unit at the same time during the second pass. The participants then continued reading and refixated the AOI or any troublesome unit during the third (error recovery) pass if any discrepancies in meaning were found.

This three-stage model was on the whole verified by the eye movement patterns of participants with a high level of predicting accuracy. In addition, the average duration of fixations denoted that the reading behavior in the first pass resembled that of normal silent reading. The average translation rates were 64 and 62 words per minute for the experts and the amateurs, respectively, reflecting around two to three times of the duration spent on reading. Finally, it seemed that expertise did not affect the translation speed, but the quality of the product. As enlightening as this study may be, there are some obvious deficits that deserve our attention. One is that participants are too few to confidently generalize the study's findings; another involves the statement by the authors: "Individual differences in expertise are reflected in the quality rather than the duration of the translation process" (McDonald & Carpenter, 1981, p. 246) Since the whole paper dealt only with the process of sight translation instead of product itself, it is too big a claim to assert the difference lies solely in quality.

Macizo and Bajo (2004) designed two experiments to examine whether the comprehension process of reading for repetition is identical to that of reading for translation, and direction of translation was also included to identify any possible facilitation effect. As mentioned before, in each experiment, eight professional translators who did not participate in the other experiment were recruited to either read to comprehend or read to translate 96 sentences. The sentence structure consisted of a relative clause and a main clause. In general, the duration of reading for translation was significantly longer than that of reading for repetition. In terms of translation from L2 into L1, this effect was only observable in the ending area of the relative clause. In addition, when reading L1 texts, participants tended to utilize pragmatic cues to help solve the ambiguities of assigning roles to each noun in the stimuli. Hence, comprehension was facilitated when the pragmatic cue was attached to the verb in the main clause. On the other hand, when reading L2 text, the pragmatic cue did not seem to be at work. But most importantly, comprehension was not negatively affected in any condition.

Macizo and Bajo (2004) spoke in favor of the horizontal view of sight translation, suggesting that, due to the prolonged reading time manifested in reading for translation, there must have been some processes, such as reformulation, working at the same time other than comprehending the message. One thing that needs to be kept in mind is that this study presented sentences word by word on the computer screen, which is a far cry from the authentic situation when a reader or translator has the freedom to go over the sentence as many times as desired without fearing that words will disappear and in turn has to try harder than usual to memorize everything. Further, the division of AOIs seems unequal. Hence, using each AOI as a basic unit for comparisons runs the risk of comparing data on different bases. Lastly, the declaration of a horizontal view is too bold a claim to make, since the definition for a horizontal/vertical process is so vague that it applies equally to a single word and a whole sentence. Owing to the fact that only one global index (reading time) was used, we have no way of confirming what really happened during the process of sight translation.

A few other studies were concerned with the cognitive load one might bear during sight translation. Alves, Pagano, and da Silva (2011) invited six translation students and six professional translators on board for an experiment investigating whether reading for different purposes impose different levels of cognitive load, reflected in the indices of reading time, the number of fixations, and fixation duration.

The experiment was divided into three modes, including reading for 1) comprehension, 2) giving summary, and 3) sight translation. All would appear in two conditions, the first of which consisted of three texts with divergent rhetorical structures on the same piece of news (193 words on average), and the second was comprised of three texts on different topics with the same rhetorical structure (105 words on average). The results of condition one showed that the total reading time for summary was the shortest, followed by reading for comprehension, and sight translation was the most time-consuming task. In terms of the number of fixations, three modes were roughly the same, while that of summary was significantly fewer than that of sight translation. As for average fixation duration, only sight translation showed significantly longer duration than the summary mode. Switching to condition two, the reading time of the task for comprehension was the shortest, followed by summary and then sight translation. When it comes to the number of fixations, only reading for summary showed significantly more fixations than that of sight translation; conversely, the average fixation duration of reading for summary was significantly shorter than that of sight translation.

Generally, the two conditions manifest the same trend. The results indicate that sight translation may carry a heavier cognitive load, reflected in significantly longer fixation duration on average and longer task time in total. Nevertheless, the texts chosen varied widely in length across the two conditions, not to mention the genre and the content, and it is risky to directly compare the figures between conditions. In addition, even though the rhetorical structures of three texts in condition one were manipulated, learning effects might still come into play because they were all about the same news; sticking each text with a certain task throughout the whole study and simply altering the task sequence also threatened the validity and reliability of the data. Last but not least, with the already few number of participants, deleting data from two of them might indeed largely twisted the trend emerging from the results.

Shreve et al. (2010) aimed to verify if sight translation is more sensitive to visual interference than written translation through manipulation of syntactic properties. Whether reading for sight translation takes up more cognitive resources than reading for comprehension was also examined. Two Spanish texts (A and B, both around 167 words) were used for sight translation, each consisting of a syntactically complex paragraph and a syntactically non-complex paragraph. Participants' eye movement data were compared with those from a bilingual reading task. In general, the authors found that there were more and longer fixations and more regressions during sight translation than in bilingual reading. Unfortunately, no inferential analysis was performed, so we don't know if the two types of tasks are fundamentally different, except that the number of regressions in sight translation was almost four times more than that in bilingual reading.

Contrasting reading measures of sight translation from syntactically complex and those from non-complex paragraphs, the authors concluded that text A showed expected syntactic effects. That is, in text A, the complex paragraph led to greater disruption than the non-complex paragraph. However, statistical analysis proved otherwise. None of the eye movement figures showed statistical difference between paragraphs with different syntactic manipulation, as did pupil size. Furthermore, for text B, in which the first paragraph contained syntactically complex sentences, eye movement data showed that processing of the easier paragraph was even more effortful. Moving onto the written translation task, the procedure was identical, and the task followed sight translation for every participant. This time, all eye movement data again showed no significant difference between conditions.

The results provided by this study suggest that, firstly, reading for sight translation is not the same as bilingual reading; secondly, the amount of cognitive resource used does not vary as a function of syntactic complexity, and cognitive load does not differ between sight translation and written translation. However, a few factors may have led the authors to this precarious interpretation. To begin with, comprehension level was not tested, and it makes little sense to analyze one's data if s/he can't even understand the content. Secondly, the text used in another bilingual reading task was of 104 words, largely different from the texts of 167 words for sight translation. Adjusting the data based on the number of words as the authors did carries the assumption that all the data may only increase or decrease proportionately. Further, although the authors contrasted the first paragraph of text A (non-complex) and that of text B (complex) and found significant difference, between-text difficulty was not controlled, so that reliable gap might have come from the disparity between texts, not syntactic complexity. Needless to say, the fixed sequence of text A preceding text B may unfaithfully tilt the results. Above all, syntactic complexity was not operationally defined, leaving all interpretation on a shaky basis.

Trying to understand sight translation more from another aspect, Chmiel and Mazur (2013) aimed at examining the influence of training on cognitive load manifested in sight translation by interpreting trainees. Eighteen participants were recruited, among which ten had received one year of interpreting training and eight for two years. Variables manipulated in the study included sentence structure (simple SVO and complex non-SVO) and frequency, for which a few low-frequency words were selected. It was assumed that participants with more training might endure less cognitive load while performing sight translation, reflected through less fixation counts, shorter fixation duration, and less total time on the whole, and also less in the number and duration of fixations when processing low-frequency words. The results indicated that one year of training might not be sufficient enough to reflect a significant difference between two groups of participants on all the indicators adopted by this study, including fixation counts, fixation duration, and total time.

Interestingly, in view of the reading pattern, it seems that simple sentences exerted more cognitive load than complex sentences, with more and longer fixations for the former. Since this finding was in stark contrast with previous studies and intuitive expectation, the authors replaced complexity with readability as the indicator, defined in terms of number of clauses per sentence and word length. This time, more readable sentences (complex non-SVO sentences in this research) showed less counts and length in fixation, suggesting that readability might be a more adequate index for examining cognitive load. The authors speculated that the non-significant findings between two groups of participants might come from the fact that less experienced participants were indeed more competent, proven by better scores on their final exam. However, the final exam of a specific course does not necessarily provide adequate evaluation of individuals' competence, since each exam may be based on different purposes and standards. This again tells us the importance of clear standards for participant selection. Besides, this study focuses itself only on the perception stage, and therefore it is plausible that the real difference might become obvious once we take final product into account.

Still another study looked into an issue that is particularly related to the industry. Dragsted and Hansen (2009) put sight translation under test to see if it could boost translators' working speed and maintain acceptable quality at the same time. The authors recruited four interpreters and four translators with at least ten years of experience. The test material amounted to 1289 words, which is probably too long and is likely to lead quickly to fatigue when coupled with an eye tracker, especially for a translation assignment. As indicated by the results, the quality of the sight-translated output was not unanimously worse than written translation. Also worthy of consideration is the fact that the reading patterns were quite dissimilar between sight translation and written translation. Focusing on sight translation performed by interpreters and translators, one may easily notice the different patterns shown by the two groups; while interpreters generally proceeded with smoother patterns, eye movements of translators were more hesitant and quite often disrupted. Quality comparison between the two groups showed immense deviations, which even held true among within-group individuals. Yet, a lack of statistical analysis strips us off the ability to conclude how each group did or if any reliable difference exists.

Sight translation as a topic has already been explored with various kinds of research tools, though in total this area still constitutes only the tip of the iceberg in interpreting studies. While most studies have been trying to infer what happens during the process through analyzing oral output, some non-traditional studies have taken a different path. This is worth celebrating because understanding how sight translation is completed is as crucial as what the final product looks like, if not more. Notwithstanding, a few reminders are in order. To begin with, many of the studies did not control variances inherent in the materials, and sometimes the sequence for stimuli presentation was not balanced, either, leading to a high possibility of interference from learning effects and biased results. Further, the number of participants or the amount of data was severely insufficient for most papers to observe certain trends or even make inferences, and criteria for selecting participants (e.g. language competence) were seldom established, so from time to time within-group variance was even more overwhelming than inter-group difference.

A related issue is that some studies recruited translators to perform sight translation, a task that is also practiced by some translators but almost never under time constraint or without a chance to modify one's own output. In this way, the observed phenomena might deviate from those of real prospective sight translation practitioners. Thirdly, descriptive statistics was still the norm, and thus we as readers could hardly tell whether the disparities between various tasks show up by chance or indeed emerge as a function of the nature of those tasks. One last observation is that, as mentioned previously, product-oriented studies can only speculate the reasons behind certain phenomena based on subjective judgment. Recently, more research methods and tools from other disciplines have joined in this line of research on sight translation, and we started to see fruitful results coming from advanced tools such as eye trackers.

2.8 Chinese-English sight translation research

In light of the need to reveal more about the processing of information during sight translation, some theses on the postgraduate level start to yield enlightening results and instill momentum into the field of translation and interpreting research. For instance, Huang (2011) recruited eighteen graduate school students of interpreting who had already finished the sight translation course in the first year at school. Based on her findings, novice interpreters performing sight translation tended to simply read the source text ahead of oral production of the target text, showing that comprehension and reformulation might come contiguously in two stages. In addition, the strategy of "reading ahead" was utilized by participants. In around three-quarters of the time, interpreting students read a certain segment of the source text while simultaneously sight-translating the previous segment. Aside from the strategy, reading for sight translation indeed resembled reading for comprehension. However, while the two tasks were similar in the first pass of reading, differences set in in later stages.

Huang's (2011) findings were partially replicated by Chen's (2013) thesis, in which experienced interpreters became the target participants. Eighteen professionals with more than 150 days of work experience were recruited. First-pass indices in general showed that reading aloud was more effortful than silent reading and sight translation. However, Chen found that there were already significant disparities between silent reading and sight translation regarding first pass indices, exhibiting a somewhat different cognitive process early in reading for the latter task. Overall, experienced interpreters spent roughly similar amount of time (and actually more, numerically) compared to interpreting students in all indices, such as first fixation duration, single fixation duration, gaze duration, rereading time, and total viewing time. But interestingly, it seems that experienced interpreters did more than comprehension during first-pass reading, a factor that could possibly explain better output quality than novices (see Chen, 2013).

Su (2013) used data from Huang (2011) and investigated what interpreters were doing during pauses. The findings suggest that, during hesitation pauses, where pauses did not correspond to adequate "timeout" intervals in the text, participants tended to look back for previous information in order to repair their understanding or output; on the other hand, during juncture pauses, where pauses were inserted right on syntactic or sentence boundaries, novice interpreters were inclined to move their eyes further down the sentence. On the whole, Su's (2013) findings corroborated Huang's (2011) claim of interpreters' reading ahead before entering the reformulation stage.

Hsieh (2014), with a comparison of eye movements during pauses between novice and experienced interpreters, reported that experience led to divergence in reading: experienced interpreters sometimes seemed to be able to offer oral renditions at the first sight of the text. That is, reading and reformulation was to a large degree completed during the first pass (also see Chen 2013), and oral output could even possibly be accomplished right off the bat as eyes move along the lines; on the other hand, novices seemed to adopt a regular silent reading approach and started reformulating after comprehension was completed. In addition, novices produced significantly more hesitation pauses than experienced interpreters, and even within the novice group, hesitation pauses were significantly more than juncture pauses, unlike experienced interpreters, for which the two types of pauses were evenly distributed.

Unfortunately, limited by time and other factors, Huang (2011) only examined the beginning area of each sentence, and therefore cannot give us the whole picture regarding the concept of "reading head". What's more, the definition of this term is still elusive. Technically, sight translation is not possible if "reading ahead" does not exist, because there is nothing to process and reformulate before one receives information. Therefore, a more proper question should be how far one reads ahead. On the other hand, Hsieh (2014), restricted by her own topic, can only show where interpreters look at during pauses. Therefore, the whole picture of reading during sight translation and also the influence of training and experience on the reading pattern remain a big puzzle to be solved.

2.9 Summary

In this chapter, we have introduced sight translation as a sophisticated activity with practical utility. Through Gile's (1997) Effort Models and relevant discussion, we can now assert that sight translation, requiring instantaneous reading and oral rendition, is unique in its own way and is by no means easier than simultaneous and consecutive interpreting.

We also reviewed literature in reading, knowing that this skill is complex in nature but rather quick in relation to time. From word identification to meaning integration, phonological, morphological, semantic, and syntactic properties all play a role in deciding the ultimate comprehension of messages conveyed through written symbols. What complicates the whole thing is that both language systems of a bilingual will likely be activated during reading, and it sometimes leads to interference, especially for highly distinctive language pairs.

Due to highly disparate syntactic features of Chinese and English, manifested through opposite principal branching directions on the phrase and sentence level, we speculate that sight translation between Chinese and English may be quite a huge challenge. And it indeed is. As shown by previous studies, expression errors emerge frequently during Chinese-English sight translation. On some occasions, participants even sacrifice accuracy under the pressure of the need for natural and fluent delivery.

Studies on sight translation have repeatedly proven that this mode of interpreting is never as easy as a mere combination of reading and production. However, with most experiments or observations solely relying on product analysis and debriefing, we only see *what* the results are. *How* sight translation is performed remains unknown, and is subject to speculation. To really understand the nature of this skill, how it is done, and even how training and experience affects the process, we need an advanced research tool that provides real-time documentation of information reception. Therefore, An eye tracker will be the best fit.

Four M.A. theses on English-Chinese sight translation have offered some great insights. First of all, it seems that reading for sight translation does resemble silent reading in large part for interpreting students, but not for experienced interpreters. Numerically, experienced interpreters spend more time than interpreting trainees in the first pass, but in the meantime the former accomplish more things, reflected through better output and the fact that they sometimes finish rendition without second-pass reading. In addition, interpreters do read ahead during juncture pauses. However, all these M.A. theses chose Chinese-to-English sight translation as their focus. Furthermore, the phenomenon of "reading ahead" was only investigated for the beginning area of each sentence or during pauses. How one works through the text and moment-to-moment changes of their fixation location generally fell outside of the research scope. Finally, though the influence of experience was partially touched upon, how training impacts the whole cognitive process continues to be blurred.

Accordingly, the current dissertation looked into English-to-Chinese sight translation, hoping to provide observations from the opposite working direction. Instead of general texts, three texts in the field of international politics by the same speaker were selected. International politics is a topic frequently encountered for interpreters, and therefore provides authenticity when examining practitioners' cognitive behavior. Besides, untrained participants with comparable linguistic competence were invited, in addition to trainee and experienced interpreters. In this way, both the impacts of experience and training might be unveiled at the same time.

This dissertation aimed to offer both “macro” and “micro” perspectives on sight translation, and specific research questions of this study are listed as follows:

1. Does reading purpose influence reading behavior (reading pattern)?
2. How does training influence sight translation?
3. How does experience influence sight translation?

According to the research questions in this study, papers targeting sight translation while also touching upon similar questions are tabulated below to serve as a reference.

Table 1 Studies addressing questions relevant to this study

Name (Year)	Participant	Goal	Task	Finding
		Eye movement research		
McDonald & Carpenter (1981)	*2 translators *2 bilinguals	*Reading pattern *ST process	*Reading for comprehension *Reading for ST	*Pattern: similar *ST process: comprehension in 1st pass; reformulation in 2nd pass; error recovery in 3rd and later passes *Speed: comprehension > ST *Experience: affecting quality, not speed
Shreve, Schäffner, Danks, & Griffin (1993)	*10 trainees *10 English majors *10 psychology majors	Reading pattern	*Reading for comprehension *Reading for paraphrasing *Reading for ST	Pattern: similar (generally, though ST took longer time)

Table 1 Studies addressing questions relevant to this study (continued)

Name (Year)	Participant	Goal	Task	Finding
Eye movement research				
Macizo & Bajo (2004)	8 translators	Reading pattern	*Reading for repetition *Reading for translation	*Pattern: dissimilar *Total time: translation > repetition (only observable at the end of the relative clause in L2-L1 translation)
Macizo & Bajo (2006)	16 translators	Reading pattern	*Reading for repetition *Reading for translation	*Pattern: dissimilar *Ambiguity and cognate facilitation: only in translation
Göpferich, Jakobsen, & Mees (2008)	*6 translators *6 trainees	Reading pattern	*Reading for comprehension *Reading for translation *Reading while ST *Reading while typing translation	*Pattern: dissimilar *Mean fixation duration: similar *Number of fixations: different (sig)
Shreve, Lacruz & Angelone (2010)	*7 1st-year trainees *4 2nd-year trainees	*Reading pattern *Visual interference *Impact of syntactic complexity	*Silent reading *ST *Translation	*Pattern: dissimilar *ST: more, longer fixations & significantly more regressions (more rereading, compared with bilingual reading). *Visual interference: similar *Syntactic complexity: no effect

Table 1 Studies addressing questions relevant to this study (continued)

Name (Year)	Participant	Goal	Task	Finding
Eye movement research				
Alves, Pagano & da Silva (2011)	*6 trainees *6 translators	Reading pattern	*Reading for comprehension *Reading for summary *Reading for ST	*Condition one (same topic with varied rhetorical structures): 1. Total time: summary < comprehension < ST (statistical analysis not available) 2. Number of fixations: summary > ST (sig) 3. Mean fixation duration: summary < ST (sig) Condition 2(varied topic with same rhetorical structure): 1. Total time: comprehension < summary < ST (statistical analysis not available) 2. Number of fixations: summary > ST (sig) 3. Mean fixation duration: summary < ST (sig)
Chmiel & Mazur (2013)	*8 2nd-year trainees *8 trainees with 2 years of training	Impact of training	ST	Number of fixations, mean fixation duration & total time: similar
Huang (2011)	*18 trainees	*Reading pattern *Reading-ahead strategy *Process of sight translation	SR RA ST	*Pattern: similar in the first pass *Reading-ahead: 72.8% (in beginning areas of sentences) *Process: vertical processing (comprehension & reformulation appearing in two stages)

Table 1 Studies addressing questions relevant to this study (continued)

Name (Year)	Participant	Goal	Task	Finding
Eye movement research				
Chen (2013)	*18 interpreters	*Reading pattern *Impact of experience *Process of sight translation	SR RA ST	*Pattern: dissimilar (already different in 1st pass, meaning additional processing for interpreters) *Accuracy: interpreters > trainees (sig) *Fluency: interpreters > trainees (sig) *Process: horizontal processing (comprehension & reformulation appearing almost concurrently)
Su (2013)	*18 trainees *18 interpreters	*Cognitive process during pauses *Juncture pause vs. hesitation pause	ST	*Three-stage processing: comprehension, reformulation, and error correction *Juncture pause: reading ahead for new information *Hesitation pause: problem solving
Hsieh (2014)	*18 trainees *18 interpreters	*Cognitive process during pauses *Juncture pause vs. hesitation pause	ST	*Three-stage processing: followed by trainees, but sometimes violated by interpreters (with production finished in 1st pass) *Number of hesitation pause: NEW > PRO (sig) *Pattern of pause: 1. Hesitation > juncture (PRO) 2. Hesitation > juncture (sig)(NEW)
Output analysis				
Viezzi (1989)	*trainees *interpreters (combined together)	Retention rate	*Listening *Reading *SI *ST	*French→Italian: ST=SI *English→Italian: ST<SI (due to structural disparity) (Statistical analysis not available for all results)

Table 1 Studies addressing questions relevant to this study (continued)

Name (Year)	Participant	Goal	Task	Finding
Output analysis				
Chiang, Kuo & Chen (2009)	*11 1st-year trainees *9 2nd-year trainees	*Impact of training *Impact of direction	ST	*Training: 1. Pause duration: 2nd-year < 1st-year (sig only for English-to-Chinese) 2. Number of pauses: 2nd-year < 1st-year (sig) *Direction: 1. Number of pauses: Chinese-to-English > English-to-Chinese (sig) 2. Pause duration: similar
Lee (2012)	*3 1st-year trainees *3 2nd-year trainees *3 interpreters	*Impact of training *Impact of expertise	ST	*Total time: experts < trainees *Major errors: experts < trainees *Minor errors: similar *Expression error: experts < trainees (Statistical analysis not available for all results)

*Note 1: SR = Silent reading; RA = Reading aloud; ST = Sight translation; SI = Simultaneous interpreting

*Note 2: PRO = Experienced interpreters; NEW = Interpreting students

As seen in Table 1 above, eye-tracking studies involving sight translation mostly focused on whether reading patterns differ between various reading activities. While a few scholars contended that reading behavior remained similar across tasks, a much higher proportion of researchers sided with the claim that the reading purpose matters. In general, studies concurred that reading for sight translation requires much more time and effort than reading for comprehension, paraphrasing, or summary. Some scholars found differences in the mean fixation duration, while others in the total number of fixations or regressions. However, local indices were seldom used to examine different stages of reading in sight translation, except for Huang (2011), Chen (2013), Su (2013), and Hsieh (2014), in which reading purposes appeared to affect only some, but not all stages. As to output analysis, the focus concentrated on the impacts of training and experience on the quality and pausing behavior. Participants with a longer period of training tended to have fewer and shorter pauses, and more experience brings better quality, reflected through less total time and fewer major or expression errors.

This study is the first to look into English-to-Chinese sight translation with an eye tracker to show the efforts required for sight translation and how it is performed by participants of divergent backgrounds. Micro-indices of eye movement that reflect cognitive load were interwoven with analyses of such macro-indices as total time, translation speed, pausing, and quality, so observations of the detail could enrich our explanation or extrapolation of more holistic data and vice versa, giving us more clues about what has actually been done in the process and the effects of different behaviors on oral production.

Chapter 3 Research Method

Sight translation involves reading for information reception and speaking for production. While this mode of interpreting resembles silent reading in the way it receives messages, it is also similar to reading aloud because both reading and oral production are required, though reformulation in another language isn't. Since silent reading and reading aloud differ from sight translation in some aspects, respectively, this study included all three tasks for comparison, in order to reveal the cognitive activities that differentiate the three tasks. In addition, quality of performance was measured to cross-reference reception and production data to get a fuller picture.

3.1 Participants

Three groups of participants participated in this experiment. The first group consisted of 18 experienced interpreters, each with more than 150 days of interpreting experience in the industry (AIIC, 2012) and had already completed formal interpreter training courses. The second group included 18 interpreting students that went to graduate institutes of translation and interpretation around Taiwan and had finished their sight translation training within three years by the time of the experiment, and hence with really limited work experience as an interpreter. Bilinguals that had not received training in interpreting skills but had similar linguistic competency to the other two groups were invited as the third group (both early and late bilinguals were eligible in this study; for definition, consult the Fédération des parents francophones de Colombie-Britannique, n.d.).

A certificate of linguistic ability was required for all participants, and the score should be no less than TOEFL 90 or on a comparable level, should certificates of other exams were used.

Information about this study was posted on social media platforms and passed along through connections of the author. Experienced interpreters with proof of work experience were all recruited via introduction from their colleagues and interpreter trainers, while interpreting trainees were mostly contacted by the author himself. Untrained bilinguals either expressed willingness to take part in the experiment through social media or were introduced by other participants.

All participants had normal or corrected-to-normal vision, and everyone signed an informed consent form before participation. Everyone that went through the experiment received NT$800 as remuneration, while those who failed calibration were still given NT$200 to cover their transportation.

3.2 Material

Three speeches from the same speaker, the Director-General of the World Trade Organization, were adapted to three 175-word texts. Rewriting was kept at minimum to maintain the authenticity of formal speeches in a diplomatic setting. Content-wise, all three texts were general without domain-specific terms or specifics of any work program. Only the mission of the WTO, the status quo of both the organization and the world, and the relationship between the two were mentioned. Two words that seldom appear in everyday conversation (i.e. Doha Development Agenda &

multilateralism) were explained during the introduction of the experiment to participants, and the Chinese equivalents were provided. Four untrained bilinguals with similar linguistic competence were asked to read all three texts and confirmed that the content of each was comprehensible, even without background knowledge.

There were no specific Areas of Interest (AOIs), but modifiers, including adjectival and adverbial words, phrases and clauses that took relatively distinct positions in Chinese and English sentences were marked for possible further analysis on syntactic differences in the future. A small portion of modifiers (two places in each text) that occupied similar positions in sentences in both languages would be regarded as the baseline when questions about the processing of syntactic differences were addressed. See the following sentence as an example:

> The founders of the WTO clearly saw that the well-being of habitats, societies, and economies are not separate.

In the sentence above, "clearly" occupies the same position in both English and Chinese – the adverbial modifier precedes the verb. Therefore, one does not need to translate "saw" first and then go back for "clearly". See another example with an adverbial phrasal modifier:

> Without trade, it is difficult to see how the world can provide for the additional two billion people expected by 2050.

Here the conditional adverbial phrase "without trade" takes the same position in both

English and Chinese – preceding the main clause; hence, the sequence of words does not have to be changed during reformulation to match the natural word order in Chinese. These modifiers were used as the baseline in the sense that the eye going through these regions should not exhibit a different moving pattern from regions without distinct adjectival and adverbial modifiers. In this way, they served as checkpoints to ensure no participant would show unusual reading habits.

As indicated above, all three experimental texts were each 175 words. The total number of adjectival modifiers that might pose challenges (again, according to the four independent untrained bilinguals mentioned before) for reformulation was 13, 13, and 12 for Text A, Text B, and Text C, respectively, whereas the number of adverbial modifiers marked for comparison was 10, 10, and 11. Indices provided by Microsoft Word were used to rate sentence difficulty and readability of the texts. The percentage of passive sentences was roughly the same for all three texts, and the other two indices indicated similar readability (see Table 2 for detail).

Table 2 Features of texts for formal trials

	Text A	Text B	Text C
Word count	175	175	175
Adjectival words & phrases	11	10	10
Adjectival clauses	2	3	2
Adverbial words & phrases	8	8	9
Adverbial clauses	2	2	2
Passive Sentences	11%	12%	10%
Flesch Reading Ease	44.0	53.1	39.6
Flesch-Kincaid Grade level	11.9	11.2	12.0

The same four untrained bilinguals were invited as raters to give subjective assessment of all the texts. They were instructed to rate each trial and to give a general score on the whole text from 1-7 (1 means really easy; 7 is equal to really difficult). The mean general score was 3 for Text A, 2.75 for Text B, and 3.25 for Text C, with Cronbach's α reaching .977, indicating high reliability and a similar difficulty level among different texts.

3.3 Design

Group (experienced interpreters, interpreting students, and untrained bilinguals) was used as an independent variable, so was condition (silent reading, reading aloud, and sight translation), making this experiment a 3 x 3 design. Three texts were used in the experiment, and the sequence of texts and tasks presented for each participant was counter-balanced respectively across subjects, as shown below in Table 3.

Through the rotation of texts and tasks respectively, we ruled out the influence of text-induced variations and practice effects (i.e. people were warmed up in the experimental setting for certain tasks). In addition, two comprehension questions were in place to ensure the content was sufficiently understood, which was a prerequisite before we could be confident that the data collected truly reflected the cognitive process.

Table 3 Sequence of texts and corresponding tasks (identical for three groups)

Subject	Sequence of text*	Designated task**
1	1, 2, 3	1, 2, 3
2	2, 3, 1	1, 2, 3
3	3, 1, 2	1, 2, 3
4	1, 2, 3	2, 3, 1
5	2, 3, 1	2, 3, 1
6	3, 1, 2	2, 3, 1
7	1, 2, 3	3, 1, 2
8	2, 3, 1	3, 1, 2
9	3, 1, 2	3, 1, 2
10	1, 2, 3	1, 2, 3
11	2, 3, 1	1, 2, 3
12	3, 1, 2	1, 2, 3
13	1, 2, 3	2, 3, 1
14	2, 3, 1	2, 3, 1
15	3, 1, 2	2, 3, 1
16	1, 2, 3	3, 1, 2
17	2, 3, 1	3, 1, 2
18	3, 1, 2	3, 1, 2

*Text A = 1; Text B = 2; Text C = 3

*Silent reading = 1; Reading aloud = 2; Sight translation = 3

A background questionnaire inquiring about participants' educational background, linguistic capabilities (self-perception and language competency certificates), and the frequency of use of interpreting/translation skills or language was administered online. This language background provided some refined indices for group categorization and factors that might possibly explain differences between individuals.

3.4 Procedure

Participants were welcomed into a dimly lit room and tested for their dominant eye, the movements of which were later recorded during the whole experiment. All participants were seated in front of a monitor and tested individually. Prior to the experiment, the researcher provided introduction, and explained the procedure and things worth notice, along with translations of two technical words. Participants were told that a blank screen with only a cross in the upper-left corner would appear every time they turn a page. They had to fixate on the cross, where the sentence would begin, for the stimuli to appear. This experiment included three texts and three conditions: silent reading, reading aloud, and sight translation. One text would be used for each condition, and in each condition a practice paragraph (2-3 sentences) would precede a formal text of four trials, which was then followed by two "yes/no" comprehension questions. All people were advised to proceed at one's own pace and adhere to how they would normally handle all three kinds of tasks. In addition, the researcher specifically emphasized that sentences in each trial were complete, so the focus should be trial-based, and they didn't have to rush into the next trial, fearing that there might be information they would need to successfully decipher the current trial.

Once participants went through the instructions, a few minutes were given so they could familiarize themselves with the technical words. A nine-point calibration and validation then followed to ensure fixation detection was sufficiently accurate. As the experiment started, the researcher would repeatedly remind the participants of the designated task during the practice trial to make sure they know what to do when the formal stimuli comes up. Upon finishing the formal trial, the researcher would again mention that, on the response box, the left button stood for “yes” and the right button meant “no” during the stage of comprehension check. The whole experiment took around 40 minutes, including breaks between tasks. Participants’ oral rendition was recorded for further analysis. Calibration would be performed ahead of each task, and re-calibration would be in place amid the process if necessary.

3.4.1 Reading span test

Discussion regarding the relationship between working memory and interpreting performance has been around for some time in the academia, but typically with the topic surrounding simultaneous interpreting. This study aimed to examine sight translation performance by three different groups, hence working memory was introduced as a variable that might impact performance in some way.

A reading span test followed the formal eye-tracking experiment. The test was designed based on Daneman and Carpenter’s (1980) study. Only the English version of the test was used, due to the fact that the source texts were written in English. A total of 42 sentences were collected from news or other reading span tests and were modified to 15-17 words per sentence. The last word of each sentence was 5-9 letters in length, while word frequency ranged from 25.10 to 194.94, with an average of 83.03.

The reading span test was run on MATLAB with a Macbook Air. The

Participant saw only one sentence at a time on the screen. Ten seconds was provided to read and comprehend the sentence and then to memorize the last word. The participant would then be directed to a yes/no question to ensure comprehension. What came up next was another sentence, which again required reading and word memorizing, followed by another comprehension question. After a set of two sentences, the participant would be asked to enter the two words that were memorized, and the tense and singularity/plurality had to be identical to how the words were used in the text but did not have to match the sequence of appearance. All participants started with bi-sentence sets, and one more sentence would be added to the bundle after every three rounds of successful recall. On the other hand, any wrong answer would end the test. A final score, ranging from 1 to 5, would be calculated by subtracting 0.5 from the number of sentences in one set the participant had right before the end of the test. For example, if a participant successfully made it into tri-sentence sets but made an mistake on one word recall, the whole test would stop after the third round of the tri-sentence trial and return a score of 2.5 (3 minus 0.5). The final score would be used as a proxy of one's reading span in this study.

3.4.2 Background questionnaire

A background questionnaire was administered online. Participants had the discretion to fill it out before or after the experiment at home. The questionnaire was divided into three sections. The first part inquired about basic personal information, including gender, age, the repertoire of language combination, and educational background. In addition, participants were asked to share how often each language is used in their daily lives, and how comfortable they are using them (only reading and speaking were asked, since sight translation involves these two skills). One question also regarded certificates of language competency. Further, participants answered for how long they have live continuously in an English environment, in which English is the dominant language for daily communication, and the purpose of their stay, such as immigration, business, or pursuit of academic achievements, traveling excluded.

The second part asked participants to indicate when they started to learn to read and speak Chinese and English, respectively. They also had to rate themselves regarding general linguistic competency. The last section of the questionnaire covered interpreting training and occupation. Specific questions concerned if participants have taken interpreting or sight translation courses and for how long. Lastly, a few questions on occupation were meant to know about participants' work experience and whether interpreting or translation skills are actually used after training (if there's any).

Variables such as educational history, training, work experience, and also subjective and objective indices of linguistic capabilities may all serve as criteria for grouping, providing us with opportunities to find patterns in this wealth of eye movement data from individuals with diverse backgrounds. However, background information was simply used to ensure each participant fit the selection requirements of the group to which s/he belonged in this study, since the indices used were already numerous.

3.5 Apparatus

Stimuli were presented on a computer monitor (1024 x 768 pixels), and eye movements were recorded with an Eyelink 1000 Desktop Mount eye tracker manufactured by SR Research. The sampling rate adopted by this study was 1000 samples per second. Stimuli appearing in Courier (22 pixels) were shown in the middle of the monitor one trial at a time (four trials for each text) on a grey background. The monitor was placed at a distance of approximately 70 cm from the participants.

3.6 Data analysis

Comprehension threshold was set to select data for further analysis. Data from participants that answered two out of six comprehension questions wrong were excluded. For first-pass fixation indices, only those between 80-800 milliseconds entered the stage of analysis, while limits were not set for non-first-pass counterparts.

This study included first fixation duration and gaze duration as first-pass indices and also calculated mean fixation duration to reflect early stages of cognitive processing. In the meantime, second-pass indices such as go-past time, re-reading time, and total view time were also examined to pinpoint more later-stage processing, like that of information synthesis, reformulation, or even production. Regression rate would also be taken into account as a supplement to data interpretation.

As sight translation involves the reformulation of input in one language into output in another language, the retrieval of two language systems is a must. This inevitably brings heavier burden on later stages of cognitive processing after words are identified. Therefore, intuitively speaking, non-first-pass reading measures (except gaze duration) might play a bigger role in exploring the mysterious mechanism behind sight translation.

Besides the objective indices used in this study, subjective evaluation of each participant's performance by two senior professional interpreters also comes in as supplementary data to give a more complete picture of what really happens during sight translation. The two raters were compensated for their contribution, and they followed the procedure and criteria for evaluation used by English-Chinese Translation/Interpreting Exam hosted by the Language Training and Testing Center. Accuracy and fluency were independently rated between 0-5 on each trial of the three experimental texts (12 trials in total). At the end, the raters had to re-evaluate the oral rendition as a whole and give a 0-5 score on the two criteria as well.

Previous studies mostly focused on either comprehension or production alone. For the former, we can only indicate how interpreters process information (i.e. sequence for processing or reformulation, or where they "might" have run into problems). However, any conclusion cannot possibly be ascertained, since we never know how efficient or effective their strategies are. On the other hand, those who only focus on production cannot tell us anything other than the end result, leaving everything in the process obscure. A study examining both comprehension and production rids itself of the drawbacks above. Therefore, as an example, when two participants use the same amount of time with widely disparate results, we may be able to infer that processing efficiency or effectiveness may be different. In addition, same quality of work completed during either the first pass or second pass of reading may even tell us of two sets of totally different cognitive processes or strategies.

3.7 Research questions revisited

Reading measures and patterns and also oral output were analyzed to find out 1) whether reading purpose changes reading behavior, 2) how training differentiates trained participants (especially interpreting students) from untrained bilinguals, and 3) how experience tells professionals and interpreting trainees apart.

It is plausible to assume that working direction might not have much impact on reading patterns and measures, since all participants were all proficient L2 English users. Therefore, based on previous findings, it was hypothesized that first-pass indices would not tell much of a difference between various groups of participants. On the other hand, it was assumed that, for untrained bilinguals, sight translation would likely be divided into three stages: normal reading during the first pass, reformulation and production in the second pass, followed by repetitive regressions for problem solving. At the other extreme end would be interpreting experts: A higher proportion of reformulation would be completed (or at least partly tackled) within the first pass of reading, indicating a boosted cognitive processing efficiency. That is, while the first-pass indices would generally level among different groups, what differs would be the quality of oral rendition and the amount of work accomplished within a similar first-pass time frame, which in turn would bring a downturn trend in non-first-pass indices and the total amount of time spent on production. Interpreting trainees, to a large extent, would stand in between bilinguals and experts, moving towards a more “professional” way of tackling this multitasking mode of interpreting.

Chapter 4 Results

The data presented below are divided into three main sections: 1) global data, or overall observation, including the total number of fixations and time spent on tasks, quality of sight translation performance, and total number of words uttered therein; 2) more local, word-based reading indices in different passes of reading, that is, indices for micro analysis: first fixation duration (FFD), gaze duration (GD), go-past time (GPT), rereading time (RRT), and total viewing time (TVT); and lastly, 3) comparisons of how three different groups performed sight translation, focusing on the reading behavior ahead of uttering the first word, how one proceeded from then on for each trial, and also other related features.

The statistical method used for each index below was a one-way ANOVA until pointed out otherwise.

4.1 Background information of the participants

The present study recruited 18 experienced interpreters, 18 interpreting students that had already completed sight translation training by the time of our experiment, and 18 untrained bilinguals (hereafter referred to as PRO, NEW, and BIL respectively in all the tables and figures; for brevity, silent reading, reading aloud, and sight translation will each be referred to as SR, RA, and ST in tables and figures).

All experienced interpreters scored higher than 100 on the TOEFL iBT test. Among these professional participants, 77.8% were female, and 66.7% ranged between 34-41 years old. While 88.9% held a master's degree, one with a doctorate, and one hadn't finished writing her master's thesis — whose data were later excluded from all the following analyses — all received their professional interpreter training at a graduate school of translation and interpretation for at least one year. Narrowing down the training to sight translation specifically, 11.1% of the experts reported that the skill training was not concentrated in one but scattered in a few related courses; 72.2% of the experts had a sight translation class for one semester, and the rest 16.7% for a year. All experts had Chinese as their native language and English as a foreign language. When asked about life experience in an English environment, 16.7% never lived in an English-speaking country, and almost half were abroad for 1-3 years, mainly for another master's degree. In the meantime, every one took translation jobs on the side, but with different degrees of involvement.

In terms of interpreting students — among whom no one scored lower than 100 on the TOEFL iBT test — 27.8% were near the end of their first-year training, and the rest were either second-year or third-year students. When asked about sight translation, 27.8% expressed that they were taught how to perform this mode of interpreting but only had some opportunities to practice every now and then without a dedicated course; on the other hand, 16.7% had more than one year of sight translation training, while 55.6% had a one-semester course. Among all the trainees, females were the majority, accounting for 72.2%. The average age range was somewhat younger, with 18-25 taking up 50% of the group and 44.4% between 26-33 years old. While all trainees were native speakers of Chinese, 11.1% reported that English was also their mother tongue. In regard to life in an English-speaking country, 22.2% of the trainees never lived in any place outside Taiwan, 38.9% under one year, 27.8% between 1-3 years, and the range of 7-9 years and above 9 years each accounted for 5.6%, mainly for academic purposes or working holidays. Aside from the fact that 11.1% of the trainees did not have any interpreting experience, 77.8% were concentrated in the range of 1-29 days. In addition, all 18 trainees were also translators, but 66.7% had translated no more than 200,000 words.

Participants with training unknown to the author were briefly asked about their experience with sight translation training in school as an informal interview. Though the length of training varied slightly for a few of the experts and trainees, all reported that chunking — the skill that cuts longer sentences into shorter and meaningful segments — and language-flexibility building were the central part of their training.

As for untrained bilinguals, one (5.6%) participant scored 94 on the TOEFL iBT test, and another 98, while all the rest were above 100, and hence the possible confounding factor of linguistic competency was ruled out. The percentage of females reached 83.3%. In this final group, around 55.6% were 18-25 years old, showing that the bilinguals in this study were not unanimously younger than interpreting students. Only 38.9% were still seniors in the university, and 33.3% were pursuing a more advanced degree at the time of the experiment. All the group members were native speakers of Chinese, with English as the main foreign language, and 27.8% had lived in an English-speaking country, mainly for academic purposes. As set by the prerequisite for recruitment, no one had ever received any interpreting training prior to the experiment. When inquired about work experience, 22.2% had 1-29 days of interpreting experience, and 50% had translation experience, among which two participants were professional translators.

The only background index that was subject to inferential statistical analysis was reading span. A one-way ANOVA showed no significant differences between groups, $F(2,50) = 2.112$, $MSE = 0.801$, $p = .132$. Numerically, interpreting students scored highest ($M = 2.667$; $SD = 1.085$), with experienced interpreters coming in second ($M = 2.618$; $SD = 0.857$), while bilinguals scored lowest ($M = 2.111$; $SD = 0.698$).

Overall, the data of one experienced interpreter (the one that had not finished her master's thesis at that time) were excluded due to extreme data patterns in comparison with other within-group members. The interpreter spent considerably longer time, both globally and locally, and had much more fixations than her peers; on the side, she sounded hesitant and repeatedly corrected herself during the process of sight translation and hence received much lower accuracy and fluency scores. The rejection rate was 1.85% in total.

4.2 Global indices: Overall observation

There were four trials for each task in this study, and therefore all figures reported in this section are trial-based, except for the number of words (Chinese characters, to be exact) spoken and the quality ratings.

4.2.1 Total number of fixations

The mean and the standard deviation of the trial-based number of fixations for all groups in all three tasks are summarized below in Table 4.

Table 4 Trial-based total number of fixations across groups in three tasks

	Silent reading	Reading aloud	Sight translation
PRO	62.824 (24.835)	70.500 (10.508)	116.485 (34.832)
NEW	75.861 (24.914)	76.500 (14.899)	131.764 (44.389)
BIL	88.972 (25.589)	93.125 (24.964)	239.653 (108.375)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

A quick look at Table 4 is already sufficient for us to know that all groups of participants did not differ that much in tasks for comprehension, but untrained bilinguals appeared to have read the text for much more times in sight translation.

Comparisons of total number of fixations between groups in each task

During the task of silent reading, an overall significant difference between groups was observed, $F(2,50) = 4.738$, $MSE = 631.042$, $p = .013$. Post hoc analyses indicated that experienced interpreters ($M = 62.824$) had significantly fewer fixations than bilinguals ($M = 88.972$), $p = .013$, while any other two groups did not differ from each other statistically.

When participants were asked to read aloud while comprehending the text, experienced interpreters ($M = 70.5$) completed each trial with the fewest fixations in all three groups, followed by interpreting students ($M = 76.5$) and bilinguals ($M = 93.125$). When performing inferential statistical analysis, it was found that the homogeneity of variances was violated ($p = .038$). In this case, Welch's *F*-test was adopted in the place of a traditional one-way ANOVA (Mellinger & Hanson, 2017; Wu, 2012). The result indicated an overall significant difference, $F(2,31.104) = 6.256$, $p = .005$. Again, that significant disparity came from the gap between experienced interpreters and bilinguals (p = .005); any other two groups could be considered equal in a statistical sense.

In terms of sight translation, the focus of the current study, experienced interpreters (M = 116.485) again had fewer fixations than interpreting students (M = 131.764), while bilinguals' number of fixations was the highest (M = 239.653), with an overall significant difference between groups found by Welch's F-test, $F(2,30.737) = 10.284$, $p < .001$. Interestingly, the same pattern for the other two tasks was not replicated here. The differential between experts and bilinguals and also that between trainees and bilinguals both reached significance ($p < .001$ and $p = .002$, respectively). On the other hand, experienced interpreters and interpreting students were similar in this task.

The significant difference between interpreting trainees and average bilinguals only in sight translation could partly reveal the effect of training, since the other two tasks stand for average comprehension, testing one's general linguistic ability, while sight translation adds another component: reformulation in another language. The similarity, statistically, between experts and trainees could hint that we need to look elsewhere for the influence of interpreting experience.

Comparisons of total number of fixations between tasks for each group

Turning to the three tasks in the current study, we aimed to understand how actual behavior differed between tasks for each group of participants. For experienced interpreters, Welch's F-test found an overall significant difference, $F(2,25.938)$ =

14.768, $p < .001$. While silent reading ($M = 62.824$) and reading aloud ($M = 70.5$) were similar, each varied significantly from sight translation ($M = 116.485$) (both *p*s < .001).

The results for interpreting students completely replicated those for experienced interpreters. Welch's *F*-test found a main significant effect, $F(2,29.676) = 12.679$, $p < .001$, with the only non-significant gap between silent reading ($M = 75.861$) and reading aloud ($M = 76.500$). Each of these two tasks was distinct from sight translation ($M = 131.764$) at a significant level, *p*s < .001.

Untrained bilinguals exhibited the same pattern. Welch's *F*-test found that the main effect was significant, $F(2,30.737) = 16.232$, $p < .001$. For silent reading ($M = 88.972$), reading aloud ($M = 93.125$), and sight translation ($M = 239.653$), the former two showed no statistical disparity, while both were distinguished from sight translation with *p*-value < .001.

4.2.2 Total time

The mean and the standard deviation of the trial-based total time for all groups in all three tasks are summarized below in Table 5.

Table 5 Trial-based total time across groups in three tasks (unit: second)

	Silent reading	Reading aloud	Sight translation
PRO	17.507 (7.418)	21.537 (2.713)	36.533 (9.315)
NEW	21.471 (7.216)	23.932 (3.946)	42.112 (11.102)
BIL	25.001 (7.217)	26.920 (6.199)	76.072 (32.731)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

As is the case for total number of fixations, total time spent by all three groups in the two comprehension tasks seemed to be much more similar than in sight translation, where untrained bilinguals required much longer time to complete the task.

Comparisons of total time between groups in each task

In terms of the average time length needed by each group of participants to finish each trial during silent reading, an overall significant effect emerged, $F(2,50) = 4.631$, $MSE = 53.02$, $p = .014$, sustained solely by the gap between experienced interpreters ($M = 17.507$) and bilinguals ($M = 25.001$), $p = .014$. Statistical analysis failed to show a significant gap between interpreters and students ($M = 21.471$), or between students and bilinguals.

Experienced interpreters ($M = 21.537$) did not differ from interpreting students ($M = 23.932$) during reading aloud, and neither did interpreting students from bilinguals ($M = 26.920$). The only significant difference existed between interpreters and bilinguals ($p = .004$), accounting for the overall significant effect found, $F(2,50) = 6.154$, $MSE = 20.713$, $p = .004$.

Sight translation tells another story. The homogeneity of variances was violated, and Welch's *F*-test found a significant main effect, $F(2,30.651) = 12.094$, $p < .001$. Experienced interpreters ($M = 36.533$) in this task spent significantly less time than bilinguals ($M = 76.072$), $p < .001$, while interpreting students ($M = 42.112$) were similar to interpreters but varied distinctly when compared with bilinguals, $p = .001$.

The between-group pattern for total number of fixations was identical to that for total time. It makes perfect sense, since more fixations inevitably lead to longer time.

However, what's important is that interpreting students, identical to bilinguals in silent reading and reading aloud so far, differed consistently from bilinguals in sight translation. If the two groups are on par in all but sight translation regarding local reading indices again, the impact of training will be even more self-evident.

Comparisons of total time between tasks for each group

Looking at experienced interpreters for cross-condition comparisons first, we found a significant main effect with Welch's *F*-test, $F(2,25.242) = 23.412$, $p < .001$. Silent reading ($M = 17.507$) was statistically similar to reading aloud ($M = 21.537$), while each was significantly different from sight translation ($M = 36.533$; *p*s < .001 for both).

The same pattern held true for interpreting students as well. Welch's *F*-test again found a significant main effect, $F(2,29.36) = 23.657$, $p < .001$. Silent reading ($M = 21.471$) was similar to reading aloud ($M = 23.932$), when the two both significantly diverged from sight translation ($M = 42.112$), $p < .001$ each.

Welch's *F*-test returned a significant overall effect for untrained bilinguals, $F(2,30.398) = 20.467$, $p < .001$. Silent reading ($M = 25.001$) was almost the same as reading aloud ($M = 26.920$), and sight translation ($M = 76.072$) obviously stood out from the rest, both *p*s < .001.

From the global data above we found a high degree of similarity between silent reading and reading aloud, indicating that the two tasks generally required a comparable amount of effort, with the additional task of vocalization for reading aloud exerting only a negligible burden. On the contrary, reformulation in sight translation demanded more cognitive efforts from the participants.

In regard to the total number of fixations and total time, experienced interpreters and interpreting students were statistically identical on both fronts. Bear in mind that, since the two groups have both received sight translation training, the main difference here is experience, as far as this study is concerned. For the time being, we haven't seen the impact of experience.

Turning to the pair of interpreting students and untrained bilinguals, the main difference is sight translation training, since bilinguals had to first have a similar level of linguistic competence to participate in the experiment. And we clearly saw the impact of training. The two groups were identical in total number of fixations and total time during silent reading and reading aloud; on the contrary, interpreting students were significantly faster than bilinguals and resorted to fewer fixations in total when it comes to sight translation. Considering the above two indices alone, it appears that training had an immediate effect on speed.

4.2.3 Quality of sight translation output

In Table 6, we first see how many words were used by each group of participants to complete the task of sight translation. The mean and the standard deviation of the total verbal output for all groups are summarized below.

Table 6 Total verbal output of three groups (word count)

	N	Mean	SD
PRO	17	364	36
NEW	18	350	29
BIL	18	370	87

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

The homogeneity of variances was violated (p = .001), so Welch's F-test was adopted, but it failed to find an overall significant difference, $F(2,30.491) < 1$, p = .390. This shows that experienced interpreters (M = 364), interpreting students (M = 350) and bilinguals (M = 370) on average uttered similar number of words for the whole task. Cross-referenced with total time spent, bilinguals must either have had more pauses or longer verbal gaps[2], since their total time was significantly longer than the other two groups.

Ratings provided by two independent professional interpreters regarding the performance of sight translation are presented in Table 7 below. Reliability between the two raters was found with Intraclass Correlation Coefficient (ICC) reaching .772 at a statistically significant level, $p < .001$, indicating that the overall reliability is "good" (between 0.75 and 0.9). The 95% confidence interval ranged from .604 to .869 (moderate to almost excellent). The aforementioned estimates were calculated based on a mean-rating (k = 2), absolute-agreement, 2-way random-effects model (cf. Koo & Li, 2016).

[2] The silent interval between any two Chinese characters will be called a verbal gap in this study.

Table 7 Sight translation performance rating (overall quality: on a 1-5 scale)

	Overall quality	Accuracy	Fluency
PRO	4.375 (0.372)	4.566 (0.262)	4.184 (0.552)
NEW	3.924 (0.380)	4.028 (0.355)	3.819 (0.626)
BIL	2.868 (0.734)	3.153 (0.844)	2.583 (0.721)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Generally, Welch's *F*-test revealed a significant main effect for the overall quality, $F(2,31.875) = 29.753$, $p < .001$. Experienced interpreters ($M = 4.375$) scored consistently higher than interpreting students ($M = 3.924$), $p = .003$, albeit both groups were numerically close and had a similarly small standard deviation. Untrained bilinguals ($M = 2.868$) scored the lowest among the three groups, and the gap between bilinguals and the other two groups respectively was quite stable in a statistical sense, both *p*s < .001.

Further breaking the index of quality down into accuracy and fluency, the two components used in this study, we found a significant main difference with Welch's *F*-test for accuracy, $F(2,30.489) = 30.038$, $p < .001$. All gaps between every two groups reached significance. Experts ($M = 4.566$) on average scored the highest, taking a lead over trainees ($M = 4.028$), $p < .001$, who in turn left untrained bilinguals ($M = 3.153$) far behind, $p = .001$. The gap between experts and bilinguals was even wider, $p < .001$. Fluency-wise, a significant main effect also existed, $F(2,50) = 30.532$, $MSE = 0.408$, $p < .001$. On the whole, experts ($M = 4.184$) significantly outperformed bilinguals ($M = 2.583$) significantly, $p < .001$, as did trainees ($M = 3.819$), outshining bilinguals with *p*-value < .001.

In this regard, training seems to improve not only on the speed for sight translation, but also accuracy and (especially) fluency, since interpreting students finished quicker and scored significantly higher than average bilinguals, and even successfully caught up with experts when fluency is concerned. On the other hand, experience in the field seems to be correlated with better quality for experienced interpreters, but not a faster pace or smoother rendition, when compared with interpreting students.

4.3 Local reading indices: Micro analysis

In the previous section, we have identified a faster speed and better accuracy and fluency in general when taking into account the impact of training, while interpreting experience mainly enhanced accuracy. For this section, we try to find out if local reading indices that narrowed the focus to the average duration spent on each word would corroborate previous findings.

4.3.1 First-pass indices

First fixation duration

First fixation duration (FFD) is the foremost index we encounter in first-pass indices; it refers to the first time a participant sets his/her eye on a new word (i.e. the first fixation), reflecting processing in an early stage.

Table 8 First fixation duration across groups in three tasks (unit: millisecond)

	Silent reading	Reading aloud	Sight translation
PRO	247.528 (20.075)	288.629 (29.248)	253.501 (30.831)
NEW	262.115 (32.761)	307.539 (45.748)	254.351 (34.758)
BIL	261.149 (20.424)	284.537 (31.122)	258.325 (42.287)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Table 8 here has summarized the mean and standard deviation of FFD for all groups in all three tasks. On the face of it, interpreting students seemed to spend longer time than the other two groups, except in sight translation, but all the gaps between groups were small, almost to a negligible level.

Comparisons of first fixation duration (FFD) between groups in each task

Statistical analysis failed to show any significant difference between any two groups in silent reading. Levene's test was significant in this case, $p = .015$, so Welch's F-test was used instead and returned a non-significant main effect, $F(2,32.432) = 2.346$, $p = .112$. This finding indicates that experienced interpreters ($M = 247.528$), interpreting students ($M = 262.115$), and even untrained bilinguals ($M = 261.149$) spent similar amount of time during this stage, as far as this index is concerned.

As for the task of reading aloud, none of the gaps between any two groups reached statistically significant level either, $F(2,32.647) = 1.581$, $p = .221$, as denoted by Welch's F-test. The disparity between the mean for any two groups out of the experienced interpreters (288.629), interpreting students (307.539), and bilinguals (284.537) was random as a result.

When it comes to sight translation, statistical analysis still failed to show a consistent gap between groups, $F(2,50) < 1$, $MSE = 1322.936$, $p = .915$. In contrast to global data, in which experts and trainees both outpaced bilinguals, it turned out that FFD during sight translation for all three groups were actually on par ($M = 253.501$ for experienced interpreters, $M = 254.351$ for interpreting students, and $M = 258.325$ for bilinguals).

Number-wise, while interpreting students were the longest in silent reading and reading aloud, bilinguals instead had the longest average duration in sight translation. This can be seen as another testament to the effect of training at work. Having practiced sight translation numerous times, trainees seemed to be consciously limiting their time spent on the initial contact with new words. Interestingly, experts' FFD appeared to move in an opposite direction, increasing by a small amount. Unfortunately, the numbers of all groups were really close, so it is not possible to verify the above speculation in this study.

Comparisons of first fixation duration (FFD) between tasks for each group

For experienced interpreters, an overall significant effect did exist across tasks, $F(2,50) = 11.386$, $MSE = 736.335$, $p < .001$. Reading aloud ($M = 288.629$) apparently required more time than silent reading ($M = 247.528$), $p < .001$, and sight translation ($M = 253.501$), $p = .002$, while the latter two were similar in nature.

In terms of interpreting students, an overall effect emerged as well, $F(2,50) = 10.19$, $MSE = 1458.096$, $p < .001$. Reading aloud ($M = 307.539$) was significantly more time-consuming than silent reading (*M* = 262.115), *p* = .003, and sight translation (*M* = 254.351), *p* = .001, while the latter two tasks were roughly on par.

Bilinguals completely replicated the patterns of the other two groups. The homogeneity of variances was violated, *p* = .007, and Welch's *F*-test found a significant main effect, *F*(2,31.311) = 3.891, *p* = .031. As seen in Table 8, silent reading (*M* = 261.149) required almost the same amount of time as sight translation (*M* = 258.325), but was differentiated from reading aloud (*M* = 284.537) significantly, *p* = .032. However, sight translation and reading aloud took equal amount of time.

The initial contact with a new word in all three groups conveyed the same message: In the early stage, silent reading and sight translation were similar in nature, while reading aloud already started to tax participants. This finding came as no surprise, since the retrieval of phonetic information of words (i.e. vocalization) during reading aloud was rather immediate. Yet, in both silent reading and sight translation, comprehension had to be completed in the first place before reformulation in sight translation would follow. That said, perhaps one of the most interesting phenomena we have observed on this index is that experienced interpreters actually showed a divergent pattern: They had longer FFD in sight translation than in silent reading, while the other two groups took an opposite approach. It seems that experts might have spent more efforts at the first encounter, but the other two groups both tried continuously to move to following targets quicker than they would do under a normal reading condition (which takes comprehension as the sole purpose of reading).

Gaze duration

In the following, we turn to gaze duration (GD), of which the mean and the standard deviation for all groups in all three tasks are summarized in Table 9.

Table 9 Gaze duration across groups in three tasks (unit: millisecond)

	Silent reading	Reading aloud	Sight translation
PRO	289.141 (32.009)	353.541 (31.748)	276.463 (38.911)
NEW	321.260 (38.612)	370.025 (48.443)	287.930 (45.658)
BIL	316.006 (18.698)	360.294 (37.292)	298.759 (53.162)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

GD entails all fixations during the first time an eye lands on a new word before leaving it, and therefore reflects a fuller picture of word identification and meaning retrieval; still, this phase counts as a relatively early stage when the comprehension of a text is concerned.

Comparisons of gaze duration (GD) between groups in each task

An overall effect was significant in silent reading, $F(2,50) = 5.396$, $MSE = 953.638$, $p = .008$. Experienced interpreters was the fastest ($M = 289.141$), using significantly shorter duration than bilinguals ($M = 316.006$), $p = .045$, and interpreting students ($M = 321.26$), $p = .013$. However, the latter two groups did not differ from each other statistically.

Unlike in silent reading, there wasn't any statistical difference between groups in reading aloud, $F(2,50) < 1$, $MSE = 1593.28$, $p = .475$. Experienced interpreters took a small lead ($M = 353.541$), followed by bilinguals ($M = 360.294$) and then interpreting students ($M = 370.025$).

GD in sight translation basically sums up how well different groups of participants comprehended the experiment text during the first pass of eye movements, and therefore considered one of the focuses in this study. As laid out in Table 9, experienced interpreters (*M* = 276.463) again only had a tiny lead in comparison with interpreting students (*M* = 287.93) and bilinguals (*M* = 298.759) without an overall effect, $F(2,50) = 1.009$, $MSE = 2154.188$, $p = .372$.

The finding here suggests that word identification and meaning retrieval might not be the reason behind the significant gap of quality and speed between bilinguals and the other two groups, since neither first-pass indices showed a significant effect. This again came as no surprise because only participants that had a comparable level of language competence participated in the experiment.

If groups start to differ only after the first reading pass, as was assumed earlier in this study, then it is possible that the real difference come from non-first-pass indices that reflect meaning integration, reformulation, or even problem solving in a later stage of sight translation. Another possibility is that, due to the strategy adopted to cope with the needs of sight translation, examination of averaged local reading indices of each word alone won't reveal the truth until aggregated indices and processing patterns are considered.

Comparisons of gaze duration (GD) between tasks for each group

Statistical analysis showed an overall effect for experienced interpreters, $F(2,50) = 24.564$, $MSE = 1182.195$, $p < .001$. Experts spent similar amount of time during silent reading (*M* = 289.141) and sight translation (*M* = 276.463), while the two measures were both significantly different from that in reading aloud (*M* = 353.541), both *p*s < .001.

Interpreting students showed an identical trend. With an overall significant effect, $F(2,50) = 15.544$, $MSE = 1974.116$, $p < .001$, post hoc analyses showed that silent reading (*M* = 321.26) and sight translation (*M* = 287.93) were comparable, while each differed from reading aloud (*M* = 370.025), both significantly ($p = .007$ for silent reading; $p < .001$ for sight translation).

An overall effect was also established across tasks for bilinguals with Welch's *F*-test, $F(2,28.868) = 11.851$, $p < .001$. The gap between silent reading (*M* = 316.006) and sight translation (*M* = 298.759) didn't reach significance, but each task was significantly differentiated from reading aloud (*M* = 360.294) ($p < .001$ for silent reading and $p = .001$ for sight translation).

For all three groups, both FFD and GD showed that, during first pass, silent reading and sight translation were a lot similar (except for bilinguals' FFD, on which sight translation was also similar to reading aloud), while the retrieval and vocalization of phonetic information of words in reading aloud taxed all participants rather early.

4.3.2 Non-first-pass indices

Go-past time

Go-past time (GPT) is the initial non-first-pass index we would turn to. GPT aggregates all the time spent after first fixating a word until crossing the same word's right-hand boundary. Therefore, from this index on, we might start to see effects of later-stage processing, including meaning integration, reformulation, or leftward clue searching when encountering problems.

Table 10 Go-past time across groups in three tasks (unit: millisecond)

	Silent reading	Reading aloud	Sight translation
PRO	333.895 (44.482)	435.883 (65.965)	389.596 (188.100)
NEW	402.647 (80.205)	470.017 (82.393)	410.824 (155.822)
BIL	413.665 (66.306)	475.199 (87.369)	431.380 (178.515)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Table 10 above summarizes the mean and standard deviation of GPT for all groups in all three tasks.

Comparisons of go-past time (GPT) between groups in each task

An overall significant effect emerged after the statistical analysis of figures in silent reading were performed, $F(2,50) = 7.505$, $MSE = 4315.107$, $p = .001$. Post hoc analyses confirmed that experienced interpreters ($M = 333.895$) spent significantly less time than both interpreting students ($M = 402.647$; $p = .013$) and untrained bilinguals ($M = 413.665$; $p = .003$), while the gap between the latter two groups failed to reach a meaningful and potent enough level.

Statistical analysis of the results for reading aloud failed to find a significant overall effect, $F(2,50) = 1.256$, $MSE = 6295.932$, $p = .294$. Though experienced interpreters ($M = 435.883$) still spent much less time than interpreting students ($M = 470.017$) and bilinguals ($M = 475.199$), all the gaps still were considered to have happened by chance.

Sight translation replicated the pattern of reading aloud. The overall effect was not significant, $F(2,50) < 1$, $MSE = 30412.412$, $p = .779$. Numerically, as seen in Table 10, experienced interpreters on average spent 389.596 ms, interpreting students 410.824 ms, and bilinguals 431.38 ms. Interestingly, the gaps between groups started to widen further, as did GD when FFD is used as the baseline, though none reached significance.

It's worth pointing out that GPT is the only index so far on which untrained bilinguals consistently spent the longest time in all three tasks. This might imply a distinct feature of bilinguals as a group: When problems arose, they tended to immediately go back to previous regions for clues more often. However, to verify the speculation, we would need to find out whether the eye moved rightward or leftward every time when there was a pause, which could not be done in this study due to the limitation of time and resource.

Comparisons of go-past time (GPT) between tasks for each group

For experienced interpreters, Welch's *F*-test found that an overall effect was significant, $F(2,27.883) = 13.77$, $p < .001$. Silent reading took the least amount of time (M = 333.895) and was significantly shorter than reading aloud (M = 435.883), p < .001. Sight translation (M = 389.596) was again comparable to silent reading, but, in contrast to FFD and GD, was no longer significantly shorter than reading aloud. This indicates that some processing needs for sight translation have pushed the number higher, possibly hinting at the variation of cognitive burden on interpreters during this transition stage between pure first-pass word processing and non-first-pass clue searching or even reformulation.

Interpreting students manifested a similar pattern, but without significance. Welch's F-test was resorted to and the overall effect was only near significant, $F(2,32.334) = 3.221$, $p = .053$. While silent reading (M = 402.647) was only slightly shorter than sight translation (M = 410.824), reading aloud (M = 470.017) still demanded a longer duration.

Bilinguals reflected a highly resembling characteristic with wider gaps. A significant main effect was not found, $F(2,50) = 1.234$, $MSE = 14632.419$, $p = .300$. Silent reading (M = 413.665) was also the shortest, followed by sight translation (M = 431.38) and then reading aloud (M = 475.199).

As was previously mentioned, GPT entails leftward meaning integration and clue searching and stands at the junction of first-pass and non-first-pass processing. At this point, we saw that in all three groups, sight translation started to "move" from silent reading towards reading aloud, duration-wise. What is worthy of attention is that the gap between silent reading and sight translation — though non-significant for all three groups — was the widest for experienced interpreters. Just as what happened on FFD, the extra burden of sight translation seemed to appear comparatively earlier for experts.

Rereading time

The next sub-section turns to rereading time (RRT). RRT can be seen as an alternative way for meaning integration to GPT, since readers don't necessarily go back to previous regions for clues whenever they encounter problems. Sometimes they continue moving further down the line, hoping to piece together information to resolve issues they met earlier.

One thing we should bear in mind is that, during sight translation, RRT doesn't always stand for meaning integration or problem solving, since reformulation between two distinct languages sometimes require multiple passes of reading to first comprehend and then organize the information in an appropriate order in the target language. Therefore, longer RRT offers the possibility of hinting at a particularly thorny linguistic structure or difficult word when translation is required.

Table 11 Rereading time across groups in three tasks (unit: millisecond)

	Silent reading	Reading aloud	Sight translation
PRO	287.838 (58.456)	345.337 (69.177)	447.022 (157.507)
NEW	367.668 (86.571)	375.722 (89.542)	455.432 (173.772)
BIL	343.745 (54.298)	368.263 (64.316)	525.526 (163.279)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Table 11 lists the mean and standard deviation of RRT for all groups in all three tasks. Note that the average duration of RRT for untrained bilinguals in sight translation, even at first glance, clearly illuminates the necessity of multiple reading passes for this task since they were never trained to more efficiently tackle the original text, and therefore had to rely on their instinct: complete comprehension first, followed by reformulation.

Comparisons of rereading time (RRT) between groups in each task

For results of average RRT in silent reading, Welch's *F*-test was adopted and a significant main effect came through, $F(2,32.431) = 6.522$, $p = .004$. Further analyses demonstrated that experienced interpreters ($M = 287.838$) were significantly faster than interpreting students ($M = 367.668$), $p = .009$, and bilinguals ($M = 343.745$), $p = .017$, while trainees and bilinguals were statistically similar.

With reading aloud in question, statistical analysis came back with a non-significant main effect across groups, $F(2,50) < 1$, $MSE = 5663.829$, $p = .469$. Experienced interpreters ($M = 345.337$) spent less time than bilinguals ($M = 368.263$), whose RRT were in turn less than interpreting students ($M = 375.722$), but only numerically.

No significant main effect was found for sight translation either, $F(2,50) = 1.211$, $MSE = 27270.043$, $p = .779$. The average duration on RRT spent by experienced interpreters ($M = 447.022$) was the shortest, followed by interpreting students ($M = 455.432$) and then bilinguals ($M = 525.526$).

During this stage of processing in sight translation, the average duration of trainees started to approximate to that of experts, while bilinguals deviated from the other two groups even further, though not enough to show a significant difference. Still, the influence of training is visible when considering the numerical gap in all previous reading indices.

Comparisons of rereading time (RRT) between tasks for each group

In regard to RRTs in all three tasks for experienced interpreters, Welch's *F*-test came back significant, $F(2,29.51) = 9.022$, $p = .001$. Silent reading ($M = 287.838$) was significantly shorter than reading aloud ($M = 345.337$) and sight translation ($M = 447.022$), $p = .035$ and $p = .002$ respectively, and the latter two groups were found to be statistically similar.

For interpreting students, the main effect was determined by Welch's *F*-test to be non-significant, $F(2,32.24) = 1.857$, $p = .172$. Although silent reading ($M = 367.668$) and reading aloud ($M = 375.722$) were each 80 ms and some shorter than sight translation ($M = 455.432$), the gaps between any two groups were still all taken as random by post hoc analyses.

The pattern of bilinguals was in direct contrast to that of experienced interpreters, with an overall longer duration for each kind of task. Welch's *F*-test found that a main effect was solid across tasks, $F(2,31.108) = 9.887$, $p < .001$. In this case, silent reading ($M = 343.745$) was still on par with reading aloud ($M = 368.263$), while each of the two tasks was significantly shorter than sight translation ($M = 525.526$) ($p = .001$ for silent reading; $p = .003$ for reading aloud).

During this non-first-pass phase, we saw that silent reading and reading aloud as a task started to converge. This is readily expected, since the extra burden borne by reading aloud would normally pressure participants into spending extra time only in the first pass. In the meantime, all the additional work required by sight translation, including reformulation in another language and reorganization of information into a proper sequence, started to add to the cognitive load of the participants, lengthening the RRT in sight translation. Indeed, as seen in our data, reading aloud consistently took up the largest amount of time in earlier stages, including FFD, GD, and GPT. Though only untrained bilinguals had a wide enough gap to significantly differentiate sight translation from the other two tasks (and for experts that meaningful gap existed only between silent reading and sight translation), all three groups were still identical in that the length of RRT in sight translation easily eclipsed that in any other task.

The peculiarity shown by interpreting students could indeed be reasonably explained. For professionals, silent reading and reading aloud were quite fast (and obviously faster than interpreting students), and therefore the cognitive load of sight translation on this index was easily observed. In terms of bilinguals, although they spent roughly the same amount of time as interpreting students in silent reading and reading aloud, the processing load they felt when performing sight translation was huge enough to lead to a largely stretched RRT. As for interpreting students, the RRT in non-sight-translation tasks were already observably longer than experts, but the cognitive load they had when performing sight translation wasn’t as immense as it was felt by bilinguals — thanks to training; hence, the gap wasn't wide enough to show a significant difference between tasks.

Total viewing time

What follows are the results for the last local reading index: total viewing time (TVT). This index reflects the total time spent on a single word, regardless of the number of passes, and hence might offer an approximate figure about how much cognitive capacity each word required for various groups of participants. The mean and the standard deviation of TVT for all groups in all three tasks are summarized below in Table 12.

Table 12 Total viewing time across groups in three tasks (unit: millisecond)

	Silent reading	Reading aloud	Sight translation
PRO	354.150 (79.287)	418.740 (52.924)	510.781 (180.305)
NEW	438.665 (113.329)	461.836 (94.379)	508.047 (202.911)
BIL	426.733 (71.777)	473.812 (79.419)	603.208 (221.944)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Before statistical analysis, the gap between sight translation and each of the other two tasks was evidently the widest for untrained bilinguals. This can be seen as a rough estimate of the positive impact of training, since TVT represents the total time spent on processing information on a word basis, instead of reflecting a single phase of one's cognitive behavior amid a certain task.

Comparisons of total viewing time (TVT) between groups in each task

A significant main effect was found for TVT duration between groups in silent reading, *F*(2,50) = 4.46, *MSE* = 8130.135, *p* = .017. Further analyses concluded that the only significant gap was the one between experienced interpreters (*M* = 354.15) and interpreting students (*M* = 438.665), *p* = .028, while any other two groups did not diverge largely enough in the time spent. Though professionals were ostensibly faster than bilinguals (*M* = 426.733), the result still failed to reach a statistically meaningful level.

For the task of reading aloud, Welch's *F*-test found an overall significant effect, *F*(2,31.879) = 3.407, *p* = .046. Experienced interpreters (*M* = 418.74) again spent the least amount of time, less than both interpreting students (*M* = 461.836) and bilinguals (*M* = 473.812). However, post hoc analyses failed to show any difference between groups.

With reference to sight translation, TVT across groups failed to show a significant effect, *F*(2,50) = 1.272, *MSE* = 41150.052, *p* = .289, as did other local reading indices for this task. Experienced interpreters (*M* = 510.781) read a bit slower than interpreting students (*M* = 508.047), and the two groups here were much faster than untrained bilinguals (*M* = 603.208). Yet, no gap was wide and consistent enough to be considered statistically meaningful.

In terms of silent reading, interpreting students and bilinguals were significantly slower than experienced interpreters in all local reading indices (but not FFD for trainees, and FFD and TVT for bilinguals). On the other hand, trainees spent numerically longer but statistically similar duration to bilinguals in all reading indices, except for GPT. From the above two observations alone, we may be able to draw two inferences: first, comprehension of words in general requires trainees in this study to devote relatively more effort; secondly, while experts might have a considerable lead in general linguistic capabilities than the others, bilinguals are indeed indistinguishable from trainees.

When it comes to reading aloud, all groups were statistically similar, but interpreting students' measures were not consistently numerically longer in most indices any more (and were shorter than bilinguals' regarding GPT and TVT). Moving on to sight translation, interpreting students even beat bilinguals repeatedly. Other than the fact that untrained bilinguals were lagging behind the other two groups right from the beginning, the divide was enlarged gradually, as the focus proceeded from FFD all the way to TVT. In the meantime, trainees were closing in on experts step by step, even leading to a shorter average TVT than experts in the end. This again proves that training seems to have had a huge impact, but of course only numerically as far as the above local reading indices are concerned.

Comparisons of total viewing time (TVT) between tasks for each group

How experienced interpreters fared with regard to TVT during the three tasks? Table 12 shows that, with silent reading ($M = 354.15$) on average shorter than reading aloud ($M = 418.74$), sight translation ($M = 510.781$) took the longest time of all. Welch's F-test found a significant main effect, $F(2,28.181) = 6.833$, $p = .004$. TVT in silent reading, unsurprisingly, was significantly shorter than reading aloud ($p = .024$) and sight translation ($p = .009$). On the other hand, the gap between reading aloud and sight translation did not reach significance.

As for interpreting students, an overall significant effect was not found, just as was the case for RRT. Welch's *F*-test returned the *p*-value at .455, $F(2,31.951) < 1$. Despite of the fact that silent reading (M = 438.665) was the quickest, followed by reading aloud (M = 461.836) and then sight translation (M = 508.047), not even one gap between any two groups was significant.

Proceeding to results of bilinguals, we did find an overall significant difference across groups through Welch's *F*-test, $F(2,31.166) = 5.784$, $p = .007$. Silent reading (M = 426.733) took the lead, as usual, and differed from reading aloud (M = 473.812) only numerically. Meanwhile, reading aloud was also similar to sight translation (M = 603.208) in a statistical sense. The only significant gap stood between silent reading and sight translation, $p = .011$.

It is inferred that professional interpreters have perhaps become more efficient in extracting information (reformulation) or organizing their output through training and years of interpreting experience. Therefore, this boost of efficiency was reflected in the non-significant gap between reading aloud and sight translation in TVT. Despite bilinguals' identical pattern with professionals in this matter, this non-significant gap cannot be explained by either training or experience. Therefore, we would contend that the already long TVT in reading aloud — longest among the three groups — was the reason behind the non-significant difference between reading aloud and sight translation for bilinguals.

4.4 Behavior during the process of sight translation

This section reveals variations or similarities between groups during the process of sight translation to better understand the preferences or strategies of different groups and also the scope and scale of the impact of training and/or experience.

4.4.1 Reading-ahead/information searching span for first utterance

How many fixations did each group of participants utilize to pre-read information before beginning oral rendition during each trial of sight translation? The results are presented in Table 13. A significant main effect was found, $F(2,50) = 6.64$, $MSE = 1127.529$, $p = .003$. On average, experienced interpreters ($M = 25.882$) bore much resemblance to interpreting students ($M = 22.569$), with a non-significant gap, while each group was statistically deviated from untrained bilinguals ($M = 59.556$) ($p = .017$ for experts; $p = .007$ for interpreting students).

Table 13 How many fixations each group had before sight translation began

	N	Mean	SD
PRO	17	25.882	27.584
NEW	18	22.569	28.982
BIL	18	59.556	41.955

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

A display of the distribution in the following Figure 1[3] sends an even clearer message. Due to training, most trainees and experts were inclined to adopt the strategy of chunking, a skill taught in graduate schools of translation and interpreting

[3] All the following figures are presented simply because the widely varied distribution tells an even more compelling story about the strategy — or inclination — each group of participants displayed.

to divide longer sentences into shorter segments so as to speed up rendition and shorten interspersed periods of silence.

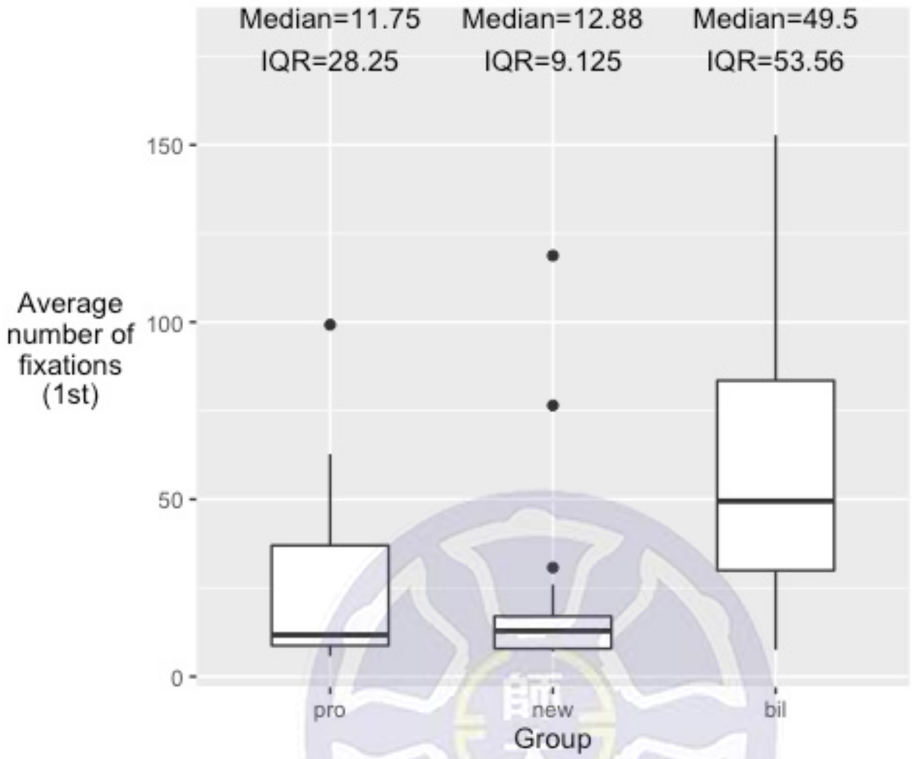


Figure 1 The distribution of how many fixations each group had before sight translation began

From the above figure, we can see that half of the experts as well as interpreting students had around 12 fixations before their oral production began (*Median* = 11.75 for the former and 12.88 for the latter), and the *IQR*[4] was 28.25 and 9.125 respectively. On the contrary, few untrained bilinguals seemed to be confident enough to just read the text for only a few times, and most felt more at ease to read through most of or the whole trial before they began interpreting (*Median* = 49.5, *IQR* = 53.56).

[4] The divide between the 3rd quantile and the 1st quantile, showing the range for the middle 50% of the samples. IQR can be a robust index when extreme values exist (Diez, Barr & Çetinkaya-Rundel, 2015).

Interpreters were often educated about not keeping the audience waiting for too long during training, hence the practice of chunking. Switching the unit for calculation from counts of fixations to time length, we can see below in Table 14 about how many seconds each group of participants waited until they started sight-translating each trial. Welch's *F*-test confirmed an overall significant effect, $F(2,32.347) = 5.376$, $p = .01$. On average, it seems that experienced interpreters ($M = 6.85$) and interpreting students ($M = 5.991$) could tolerate a similar length of silence before starting interpreting, while each group was significantly divergent from untrained bilinguals ($M = 17.271$) ($p = .017$ for experts; $p = .01$ for trainees).

Table 14 Silent period before first utterance (unit: second)

	N	Mean (s)	SD
PRO	17	6.850	7.557
NEW	18	5.991	7.803
BIL	18	17.271	12.832

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Bilinguals seemed to be more comfortable if they could first digest more information before opening their mouth for the first time. Figure 2 clearly shows the distribution of each group. The trainee group had the narrowest distribution, with *Median* standing at 3.034 and *IQR* at 2.634 seconds. Experienced interpreters were quite similar, with *Median* at 2.838 and *IQR* at 6.268 seconds. At the other extreme were the bilinguals (*Median* = 13.68; *IQR* = 14.074).

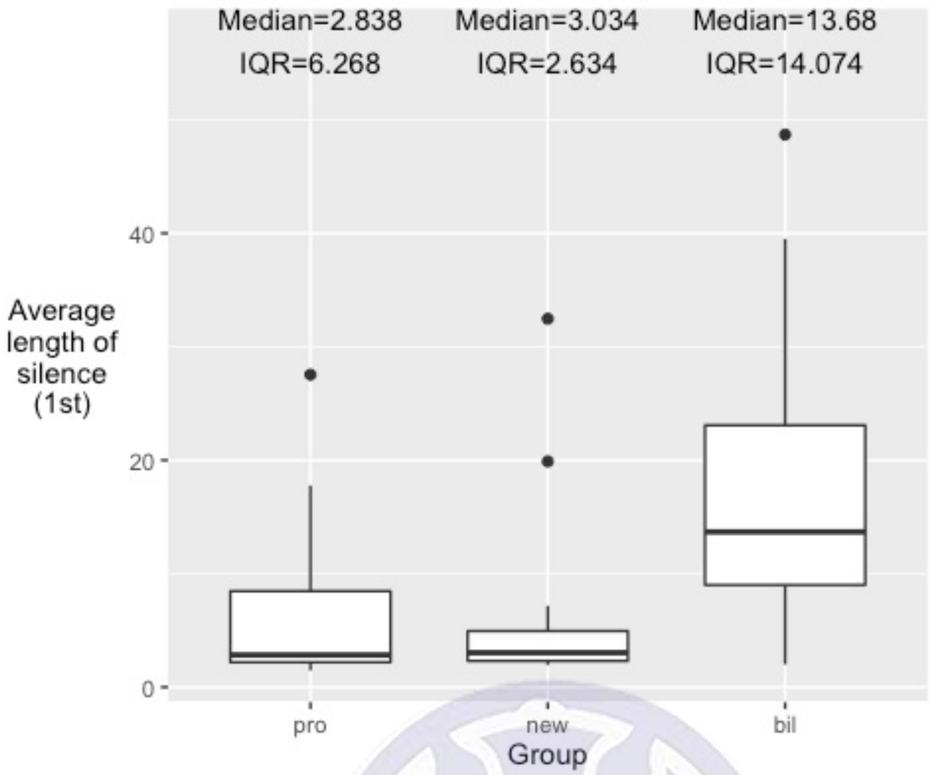


Figure 2 The distribution of how many seconds each group waited before sight translation began

Generally speaking, we can see the effect of training, which differentiated bilinguals from the other two groups, especially for interpreting students because they were very consistent in how they handled the very first moment of silence during each trial. Interestingly, thanks perhaps to experience accumulated in the profession of interpreting, some experts seemed to be willing to let the audience sit for a while and were still able to maintain the same level of fluency, while almost all trainees strictly observed the principles they were taught during training.

4.4.2 Reading-ahead/information searching span along the way

Once participants began sight-translating the text, how did they steer through the whole trial? For this section, how many times participants read the text to search for relevant information before producing each Chinese character are shown in Table 15. The first character was excluded from the analysis here because this section focuses only on what participants did to keep their oral rendition going after they started. Welch's *F*-test found an overall significant effect, $F(2,27.003) = 17.419$, $p < .001$. Further analyses showed that all the gaps between every two groups were significant. Experienced interpreters ($M = 1.501$) had the fewest average number of fixations, with the *p*-value at .002 when compared with interpreting students ($M = 1.812$) and smaller than .001 with bilinguals ($M = 2.739$). Trainees also read significantly fewer times than bilinguals, $p = .007$.

Table 15 Average number of fixations before each non-first utterance

	N	Mean	SD
PRO	17	1.501	0.143
NEW	18	1.812	0.307
BIL	18	2.739	1.096

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

The distribution of each group is provided below. As is portrayed in Figure 3, experienced interpreters this time had the narrowest distribution (*Median* = 1.531; *IQR* = 0.16), followed by interpreting trainees (*Median* = 1.827; *IQR* = 0.462). Untrained bilinguals again spread widely across the spectrum (*Median* = 2.538; *IQR* = 0.703). Training seemed to have taken effect in this non-first-utterance scenario as well, reducing the fixations needed in each interval consistently. Practical experience, on the other hand, may answer (better than it could do regarding the initial silence) how a dependence on significantly fewer fixations was possible for experts, compared with interpreting students, who also received training. Maybe experts are more resourceful, or more flexible, in finding suitable linguistic correspondences, so they wouldn't get bogged down easily by distinctive linguistic structures in this experiment.

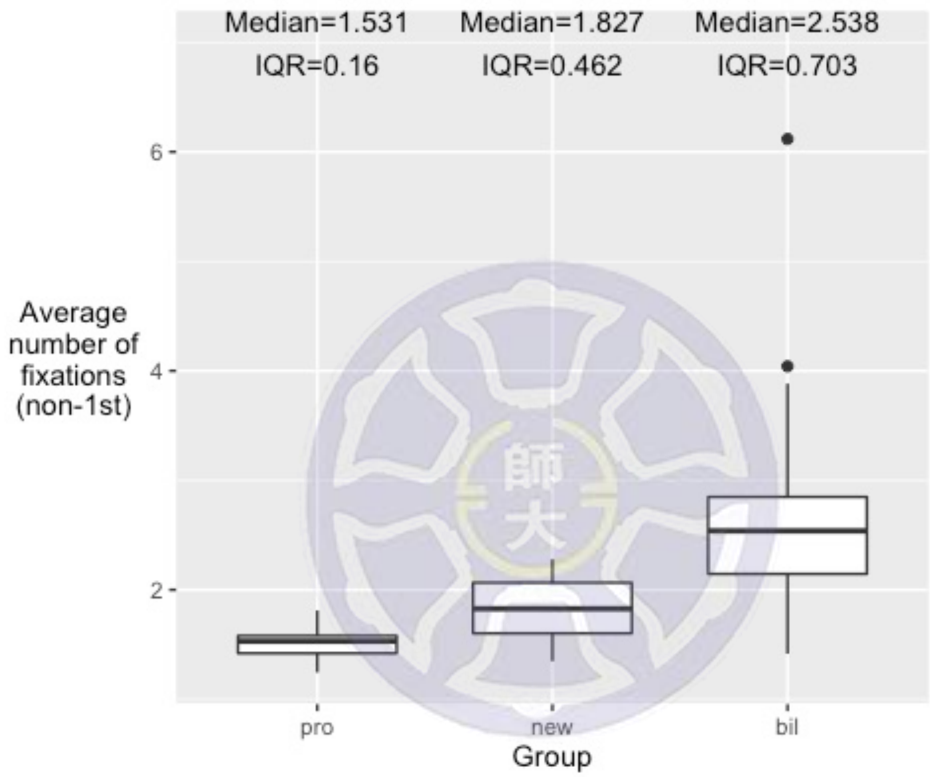


Figure 3 The distribution of how many fixations each group relied upon to keep oral rendition going after sight translation began

4.4.3 Features of all participants in sight translation across groups

In addition to total time, as well as the number of fixations for each trial, we present results in this section about other features of sight translation readily observable by the audience.

Pauses are the first feature seen below. Six native Chinese speakers were invited to mark on transcripts of participants' sight translation output all the places where they felt that the interpreter paused for a moment, long and short. Most of the time (around 90%) only verbal gaps — intervals starting from the onset of a character to the onset of the next character uttered — longer than 600 ms were marked. Therefore, 600 ms was used as the threshold to define a pause in this study. Further, pauses were divided into two categories: juncture pause and hesitation pause. A juncture pause happens right on the boundary between grammatical units, such as a sentence, a clause, or even a topic, as Chinese syntactic structures adhere to the principle of topic-comment combination. A hesitation pause indicates inappropriate stops during one's output production that are jarring to one's ear. The author alone was responsible for sorting all the pauses, but the punctuation marks used in the transcripts of the sight translation output typed by two Chinese native speakers were also taken into consideration.

Table 16 offers insight into the percentage of pauses sensed by the audience. The actual number of pauses was aggregated for each trial, and the average number of pauses across trials for each person was calculated and then divided by the total number of verbal gaps. The final figure represents the percentage of gaps observable by the audience.

Table 16 The percentage of pauses among all verbal gaps sensed by the audience across groups

	N	Mean (%)	SD
PRO	17	9.772	2.946
NEW	18	13.900	3.667
BIL	18	19.440	5.171

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

As seen in Table 16, experienced interpreters ($M = 9.772$) had a smaller percentage of observable pauses than interpreting students ($M = 13.9$) and bilinguals ($M = 19.44$). An overall effect was significant, $F(2,50) = 25.123$, $MSE = 16.44$, $p < .001$, and the difference between experts and trainees was significant ($p = .015$), so was that between each of the aforementioned groups and bilinguals ($p < .001$ for experts; $p = .001$ for trainees).

Average length of the pauses across groups and their distribution are provided in the following table. It was found that all experts and trainees had an average length of pauses under two seconds, a threshold only met by around half of the bilinguals, with the maximum average pause reaching 4.901 seconds (see Table 17). Welch's *F*-test came back with an overall significant effect, $F(2,29.558) = 14.585$, $p < .001$. Professionals' pauses ($M = 1.004$) were significantly shorter than trainees' ($M = 1.31$) and bilinguals' ($M = 2.06$), $p = .002$ and $p = .001$ each. In the meantime, the difference between the latter two groups also reached significance, $p = .016$.

Table 17 Average length and the distribution of pauses across groups (unit: second)

	Median	IQR	3rd Qu.	Max.	Mean	SD
PRO	0.944	0.228	1.115	1.474	1.004	0.196
NEW	1.268	0.361	1.499	1.796	1.310	0.285
BIL	1.852	0.72	2.168	4.901	2.060	0.996

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

After knowing the average length and distribution of pauses for each group, we wonder how much information the participants were looking for during those observable pauses. The distribution of the number of fixations each group had on average during each pause is laid out in Table 18.

Table 18 Number of fixations during each pause across groups

	Median	IQR	3rd Qu.	Max.	Mean	SD
PRO	2.955	0.808	3.522	4.833	3.149	0.651
NEW	4.069	1.708	4.670	5.919	3.985	1.109
BIL	5.414	2.162	6.750	16.020	6.361	3.234

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Experienced interpreters (M = 3.149, SD = 0.651) were quite consistent in their approach towards keeping the rendition going without pausing for too long, matching the pattern exhibited in non-first-utterance reading-ahead/information searching span. Interpreting students (M = 3.985, SD = 1.109) at times stumbled a little bit, and the highest mean number of fixations was 5.919. On the other hand, bilinguals seemed to be "fumbling" for information from time to time, with the maximum mean number reaching 16.02 fixations (M = 6.361, SD = 3.234). Welch's F-test found a significant main effect, $F(2,28.729) = 10.952$, $p < .001$. While experienced interpreters on average had significantly fewer fixations than both interpreting students ($p = .028$) and bilinguals ($p = .002$), the trainees also managed to distance themselves from the untrained, $p = .02$.

One thing worthy of attention is that the maximum number of 4.901 seconds of pause in the above Table 17 means that one member of the bilingual group on average paused for 4.901 seconds every time. The same goes for Table 18, denoting that one bilingual had on average 16.02 fixations during each pause. Seen in this way, the gap between the trained and the untrained could actually be huge.

However, a look at raw data in Table 19 about the average number and distribution of fixations during each pause across groups may point us to an interesting fact. The raw data here mean that observations regarding the number of fixations in each pause were categorized by group only, without taking the individuality of each person into consideration. In this way, the results will give us some clue about the distribution of each group, as all pauses are treated equally as items of the same feature.

As stated in Table 19 below, experts ($M = 0.809$) read fewer times than trainees ($M = 1.058$) and bilinguals ($M = 1.706$) during each pause, and the main effect remained significant, $F(2,50) = 9.295$, $p < .001$. Nonetheless, we found that all three groups were congruent in that 75% of the pauses had only one fixation (as shown by the *3rd Qu.* in the Table). It means participants showed the same pattern in 75% of the pauses, and the differences were concentrated in the rest 25%, which were the long pauses when real problems occurred.

Table 19 Average number and distribution of fixations during each pause across groups: Group-based raw data without calculating the mean for each participant

	Median	IQR	3rd Qu.	Max.	Mean	SD
PRO	1	1	1	5	0.809	0.752
NEW	1	1	1	16	1.058	1.516
BIL	1	1	1	215	1.706	8.818

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Participants with training by and large manifested comparable behavior. That is, these two groups of people did not have to or want to spend too much time reading in advance or digesting information extensively before they resumed interpreting. On the contrary, untrained bilinguals had to take in a lot more information to be ready to proceed with the sight translation task.

On the face of it, experienced interpreters didn't seem to get bogged down at all by problems encountered, with at most five fixations in a single pause. We can tell that interpreting students, reading measures and other reading behavior considered, were trying hard to "keep moving" along the way but tripped a bit when real problems arose from time to time, leading to some longer pauses and more fixations in extreme conditions. Bilinguals, on the other hand, really stopped and focused on solving the issue when one came up before continuing with the task, hence a higher percentage of much longer pauses with many more fixations.

Table 20 below gives us information on the average length of verbal gaps. As stated in previous sections, a verbal gap is the interval between the onsets of two characters. Information about the average length of verbal gaps tells us the speed — or pace — of sight translation of each participant. For this measure, only gaps within one positive and negative standard deviation were included in order to avoid the influence of extreme values, which were indeed a lot more extreme and also more frequent for untrained bilinguals. In the end, only around 69% of the gaps in the middle of the distribution were taken into consideration. In this way, if a significant difference is found between groups, we can more confidently conclude that groups do in fact vary in their average sight translation speed.

Table 20 Average length of verbal gap: How long each group of participants waited until saying the next character out loud (unit: millisecond)

	N	Mean (ms)	SD
PRO	17	231.7	24.504
NEW	18	258.9	33.898
BIL	18	345.6	103.227

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

As a result, a main effect was found significant by Welch's *F*-test, $F(2,29.974) = 12.315$, $p < .001$. The mean verbal gap of experienced interpreters ($M = 231.7$) was significantly shorter than that of interpreting students ($M = 258.9$), $p = .027$, and untrained bilinguals ($M = 345.6$), $p = .001$. The gap between trainees and bilinguals also appeared significant, $p = .008$ (see Table 20).

From the audience's point of view, experienced interpreters had substantially fewer detectable pauses than interpreting students, whose pauses were in turn fewer than participants' without training. The same pattern held true across groups when it comes to the pace of sight translation in listeners' ears, with experts' pace being the fastest and bilinguals the slowest.

4.4.4 Behavior of rereading and features in pauses

This section is comprised of data from five experienced interpreters, nine interpreting students, and seven untrained bilinguals. Due to uncontrollable head movements and other unforeseeable reasons — perhaps a reflection of the cognitive load experienced during sight translation — a little more than half participants' eye movements didn't land exactly on lines of words, though all passed initial calibration without a hitch and subsequent re-calibrations were conducted between trials when problems arose. To avoid over-speculation, a total of only 21 participants were included for further word-based examination, such as how many times one reread a word, and whether one was searching for information in a PBD or non-PBD unit.

Behavior of rereading

The mean and the standard deviation of rereading counts for all groups in all three tasks are presented below in Table 21.

Table 21 Rereading counts across groups in three tasks

	No. of people	Silent reading	Reading aloud	Sight translation
PRO	5	0.405 (0.564)	0.178 (0.093)	1.241 (0.735)
NEW	9	0.564 (0.536)	0.255 (0.221)	1.260 (0.910)
BIL	7	0.508 (0.208)	0.261 (0.139)	1.777 (1.023)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

Reading aloud seems to have the fewest rereading counts and a much smaller standard deviation in general. The need to speak out the written text perhaps offered some buffer, so the participants had slightly more time to process deeper during the first pass of reading. Sight translation, not unexpected, had much more traces of rereading. Although participants with training reread comparatively fewer times, but the differences between all three groups are not that obvious. Notwithstanding, interpretation of the results should be taken more conservatively.

Comparisons of rereading counts between groups in each task

An overall significant effect was absent across groups for rereading counts in silent reading, $F(2,18) < 1$, $MSE = 0.213$, $p = .827$. Experienced interpreters ($M = 0.405$) on average had fewer rereading counts than interpreting students ($M = 0.564$) and bilinguals ($M = 0.508$) respectively, but none of the gaps between groups reached significance.

Moving to reading aloud, experienced interpreters ($M = 0.178$) again had the lowest number of rereading counts per word, followed by interpreting students ($M = 0.255$) and bilinguals ($M = 0.261$). Again, none of the between-group gaps reached significance, $F(2,18) < 1$, $MSE = 0.03$, $p = .672$.

The pattern in sight translation was identical to that in reading aloud. Experienced interpreters ($M = 1.241$) ostensibly re-fixated words after first leaving the region fewer times than interpreting students ($M = 1.26$) and bilinguals ($M = 1.777$). Though the gap between the trained and the untrained was much larger than in the other two tasks, all gaps still happened by chance, according to inferential statistical analysis, $F(2,18) < 1$, $MSE = 0.837$, $p = .48$.

It was previously inferred that the trained participants must have had fewer passes of reading to finish sight translation with similar duration in all reading indices but significantly less total time and fewer number of fixations. The results for rereading counts, albeit failing to show a meaningful difference, are still clear with the numerical differentials. It might be that the insignificant gap made a real difference in the end when added up. Still another possibility is that trained participants did sometimes have fewer passes of reading but was not frequent enough when evened out across four trials; on the other hand, the practice of chunking helped trained participants find suitable correspondents more easily. The above two reasons, combined together, explain the significantly fewer total fixations and less total time but only numerically fewer rereading counts for trained participants. Or, it might simply be that the samples gathered for this detail analysis are not representative.

Comparisons of rereading counts between tasks for each group

For experienced interpreters, silent reading required an average of 0.405 times of rereading on each word. The number was close to and slightly higher than reading aloud ($M = 0.178$), but was largely fewer than sight translation ($M = 1.241$). There was an significant overall effect, $F(2,12) = 5.418$, $MSE = 0.289$, $p = .021$, but the effect came solely from the significant gap between reading aloud and sight translation, $p = .028$.

Interpreting students manifested the same pattern as experts. The average count in silent reading ($M = 0.564$) was a bit higher than reading aloud ($M = 0.255$) and much lower than sight translation ($M = 1.26$). Welch's F-test established a significant main effect, $F(2,12.674) = 5.741$, $p = .017$, and the only significant gap stood between reading aloud and sight translation, $p = .026$.

Bilinguals perhaps adequately demonstrated how situations would pan out for average people without training. On the face of it, reading aloud ($M = 0.261$) was again the lowest, followed by silent reading ($M = 0.508$), while sight translation ($M = 1.777$) occupied the other end of the spectrum. Yet, this time, with a significant main effect confirmed by Welch's F-test, $F(2,10.263) = 9.614$, $p = .004$, the only non-significant gap was between silent reading and reading aloud. The average rereading count of reading aloud was significantly lower than that of sight translation (also true for the other two groups), $p = .018$. What's more, silent reading significantly distanced itself from sight translation as well, $p = .038$.

Due to the subtask of vocalization required for reading aloud, eyes were compelled to linger a bit longer on words, thereby reducing the need to reread words. Hence, for all three groups, reading aloud had the lowest number of rereading counts. The narrower gap between silent reading and sight translation across all groups serves to point out one more similarity between these two tasks. However, perhaps the thing worthy of most attention is that, participants with training on average managed to narrow the gap between silent reading and sight translation to a non-significant level. This is indeed amazing, since the implication is that experts and trainees actually reduced the repetitive needs of reading for reformulation to a level that's close to pure reading.

Features in pauses

The following sub-section moves onto features in pauses. We first divided them into hesitation pause and juncture pause and then calculated their percentage. The mean of each type of pause for all groups are summarized below in Table 22.

Table 22 The percentage of hesitation pause vs. juncture pause among all the pauses made by participants across groups when performing sight translation

	N	Pause type	Mean (%)	SD
PRO	5	Hesitation pause	62.816	10.882
		Juncture pause	37.184	10.882
NEW	9	Hesitation pause	55.384	2.575
		Juncture pause	44.616	2.575
BIL	7	Hesitation pause	59.034	5.391
		Juncture pause	40.966	5.391

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

The percentage of hesitation pause was consistently higher than that of juncture pause for all groups. What may be counter-intuitive is the highest percentage of hesitation pause for experts, but one thing to bear in mind is that the figures here represent only the percentage, regardless of the actual number of pauses in each category.

Comparisons of different types of pauses between groups

In this paragraph, only results for hesitation pause were compared. We believe this will suffice, since the two types of pauses together constitute 100% for each person. According to Table 22, experienced interpreters (*M* = 62.816%) on average had higher percentage of hesitation pause than bilinguals (*M* = 59.034%), while interpreting students' (*M* = 55.384%) proportion was the lowest. Still, an overall main effect was found non-significant by Welch's *F*-test, $F(2,7.36) = 2.142$, $p = .185$.

Comparisons of different types of pauses for each group

For experienced interpreters, a dependent two-tailed *t*-test was adopted to examine whether the percentage of hesitation pause and that of juncture pause were different and found a non-significant main effect, $t(4) = 2.633$, $p = .058$. Although hesitation pauses (*M* = 62.816%) seemed overwhelmingly more than juncture pauses (*M* = 37.184%), statistical analysis determined that the gap simply showed itself by chance.

Interpreting students, in contrary to experts, turned out to have a significant gap between the two kinds of pauses. We can see that the hesitation pause (*M* = 55.384%) was generally higher in percentage than the juncture pause (*M* = 44.616%). Though the superficial pattern looked exactly like that of experts', even with a narrower gap in between, a dependent two-tailed *t*-test returned a significant difference, $t(8) = 6.272$, $p < .001$.

Bilinguals completely replicated the pattern of trainees. The hesitation pause on average accounted for 59.034% of all the pauses, while the other 40.966% were juncture pauses. A dependent two-tailed *t*-test found a significant difference between the two categories, $t(6) = 4.433$, $p = .004$.

Interpreting students' and untrained bilinguals' significantly higher percentage of hesitation pause shown in Table 22 demonstrate that the cognitive burden of sight translation was huge enough to grab their full attention and did not leave enough resources for them to tackle pauses with finesse. Therefore their struggles were faithfully reflected through hesitation pauses. On the other hand, despite the highest percentage of hesitation pause, experienced interpreters indeed had the lowest number of hesitation pauses. What's more, the percentage of each type of pause was not statistically different from the other for this group. This means while experts as a group had smooth, fluent, and accurate rendition with really few pauses, these participants were able to control themselves and pause at the right time so as not to create dissonance in the ears of the audience.

Fixations on principal branching direction (PBD) units in pauses

In previous sections it was found that there was only one fixation during 75% of the pauses for all three groups, while participants were gathering information to a different extent in the rest 25% (see Table 19). What were the participants reading in order to keep their rendition going? As we have mentioned, Chinese and English were contrastive in the principal branching direction of adjuncts. Would these adjuncts create problems in sight translation? Further, would there be any difference in the two distinctive kinds of pauses? To answer the questions above, the percentage of fixations on PBD units in pauses, and further, in hesitation pauses and juncture pauses were calculated respectively, and the mean and standard deviation for all groups are summarized below in Table 23.

Table 23 The percentage of fixations on PBD units of all fixations in pauses (percentage in hesitation pauses and juncture pauses also tabulated individually)

	N	Percentage of fixations	Mean (%)	SD
PRO	5	Fixations on PBD units in pause	62.479	7.006
		Fixations on PBD units in HP	64.547	11.570
		Fixations on PBD units in JP	58.610	15.698
NEW	9	Fixations on PBD units in pause	56.833	10.292
		Fixations on PBD units in HP	61.443	14.350
		Fixations on PBD units in JP	49.631	10.745
BIL	7	Fixations on PBD units in pause	59.001	11.058
		Fixations on PBD units in HP	61.693	15.957
		Fixations on PBD units in JP	56.772	8.186

*PBD = Principal branching direction
*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*HP = hesitation pause; JP = juncture pause

Principal branching direction (PBD) units attracted more attention in hesitation pauses than in juncture pauses: The percentage of fixations on PBD units was on average over 60% of all fixations in each hesitation pause for every group and was lower in the other type of pause.

Comparisons of percentage of fixations on principal branching direction (PBD) units in different types of pauses between groups

For experienced interpreters, on average, 62.479% of the fixations in each pause were stapled to PBD units, while it was 56.833% for interpreting students and 59.001% for bilinguals. No significant main effect was found, $F(2,18) < 1$, $MSE = 98.743$, $p = .603$.

We delved deeper to see if the percentage of fixations on PBD units in each type of pause would differ across groups. As stated in Table 23, experienced interpreters had 64.547% of their fixations on PBD units during hesitation pauses, while it was 61.443% for interpreting students, almost identical to bilinguals' 61.693%. Still, the overall effect was non-significant, $F(2,18) < 1$, $MSE = 206.136$, $p = .920$. In terms of fixations on PBD units in juncture pauses, experienced interpreters had the highest percentage ($M = 58.61\%$), followed by bilinguals ($M = 56.772\%$) and then interpreting students ($M = 49.631\%$), exhibiting the exact same pattern as in the other type of pause. Once again, the main effect did not reach significance, $F(2,18) = 1.29$, $MSE = 128.412$, $p = .299$, making all groups similar in this regard.

All the between-group comparisons in the section regarding behavior of rereading and features in pauses failed to show a significant effect, denoting that reading behavior and strategies during pauses were similar across groups. Though the percentage of hesitation pause of experts was the highest, we should keep in mind that the actual number of pauses of this group was in fact the fewest and was around only half when compared with bilinguals. On the other hand, the generally higher percentage of fixations on PBD, regardless of groups and pause types, is a testimony to the tricky feature of this kind of structural disparity in English-Chinese sight translation.

Comparisons of percentage of fixations on principal branching direction (PBD) units in different types of pauses for each group

As in Table 23, it appears that on average there was a bigger chance of fixating on PBD units in hesitation pauses (M = 64.547%) than in juncture pauses (M = 58.61%) for experienced interpreters. However, a dependent two tailed t-test failed to show a significant difference, $t(4) < 1$, $p = .585$. Interpreting students tended to fix their eyes upon PBD units more frequently in hesitation pauses (M = 61.443%) than in juncture pauses (M = 49.631%) as well. This time, a dependent two-tailed t-test found that the difference just crossed the threshold of significance level, $t(8) = 2.304$, $p = .05$. As for bilinguals, the probability of PBD units attracting the eyes was higher in hesitation pauses (M = 61.693%) than in juncture pauses (M = 56.772%) as well, but a dependent two-tailed t-test failed to find a significant difference, $t(6) = 1.227$, $p = .266$.

Hesitation pauses appear when the process of the sight translation cycle (reading-reformulation-production loop) is interrupted. Judging from the data at hand, when faced with difficulties, trainees did have a more consistent tendency to look at PBD units for clues. On the other, experts and bilinguals did not manifest this distinct pattern. But still, the percentage of fixations on PBD units in hesitation pauses on average exceeded 60% for all the groups.

The PBD percentage in juncture pauses was significantly lower for trainees. It's somewhat unexpected, since when one successfully finishes the previous segment, it's natural to pre-read tricky structures that lie ahead to ensure a smooth rendition when moving onto the next block of information. Interpreting students, as high-achieving language users, should know better that the structural differences between Chinese and English might present great difficulties that "ruin the show". Therefore, it's normal for experts and bilinguals to look at PBD units around 57% of the time in juncture pauses. As for interpreting students, we can only speculate that the training drills were so deeply ingrained in their processing habits that they adhered strictly to the skill of chunking and focused entirely on finishing sight-translating the target segment.

As a matter of fact, contrastive PBD units in this experiment never appeared right after a period (and few followed a comma except for sentential PBD units). Therefore, trainees sticking to the approach taught in training did indeed have fewer opportunities to pre-read PBD units in juncture pauses. Nonetheless, it's worth pointing out that, even in the case of interpreting students, the PBD percentage still reached almost 50%.

Comparisons of processing load between non-PBD and PBD units for each group

A generally higher percentage of fixations on PBD units in pauses prompted the comparisons of processing load between non-PBD and PBD units. Two-tailed dependent *t*-tests were used to examine whether the average fixation duration of one type of unit differed from the other regarding different indices within each group.

The results for the task of silent reading are summarized in Table 24. For experienced interpreters, almost all indices showed longer mean fixation duration on PBD units than on non-PBD units, except for RRT (M = 266.997 for the PBD vs. M = 289.181 for the non-PBD). However, statistical analyses did not find any significant difference between the two types of units on any index.

The pattern shown by experts was completely replicated by interpreting students. RRT was the only index with which the mean fixation duration on PBD units was shorter (M = 367.093 vs. M = 392.482 for non-PBD units). Dependent *t*-tests again did not find any significant difference between non-PBD and PBD units (fore brevity and clarity here, only those unique phenomena worthy of notice were supplied with statistical results. Please see Appendix V for the rest of the details concerning this whole subsection, including the actual numbers of statistical values).

Untrained bilinguals, on the other hand, consistently had longer mean duration on PBD units than on non-PBD units. What's more, a dependent *t*-test found that GD on the baseline (i.e. non-PBD units) (M = 298.983) was significantly shorter than that on PBD units (M= 327.851), $t(6) = -2.739$, $p = .034$. Baseline RRT (M = 321.147) was also significantly shorter than RRT on PBD units (M = 386.138), $t(6) = -2.715$, $p = .035$.

Table 24 Mean fixation duration on non-PBD and PBD units in silent reading across groups (unit: millisecond)

	PRO		NEW		BIL	
	Non-PBD	PBD	Non-PBD	PBD	Non-PBD	PBD
FFD	231.381	247.789	260.461	263.149	244.402	258.438
	(28.737)	(13.288)	(32.817)	(31.243)	(26.752)	(24.939)
GD	267.813	279.799	318.318	333.317	298.983	327.851
	(24.929)	(14.539)	(43.519)	(33.470)	(22.376)	(31.039)
GPT	291.228	348.196	401.115	414.750	400.942	438.554
	(37.679)	(63.148)	(89.796)	(84.377)	(83.917)	(83.388)
RRT	289.181	266.997	392.482	367.093	321.147	386.138
	(77.190)	(83.827)	(108.171)	(108.624)	(45.911)	(34.761)
TVT	351.435	356.679	442.964	454.172	402.035	446.013
	(132.405)	(117.753)	(126.945)	(135.624)	(57.074)	(60.134)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*FFD = First fixation duration; GD = Gaze duration; GPT = Go-past time; RRT = Rereading time; TVT = Total viewing time
*PBD = Principal branching direction

Compared with average words, units embedded in contrastive structures (i.e. PBD units) required bilinguals to spend significantly more time when it comes to meaning interpretation in a relatively later stage of the first pass and meaning integration in non-first passes. This group of participants exhibited a more effortful behavior when making connections between antecedents and adjuncts were constantly necessary to figure out the message conveyed. Notwithstanding, the overall time needed to comprehend words in contrastive structures did not pose a significantly larger challenge. On the contrary, processing of baseline and PBD units seemed similar for both experts and trainees.

Table 25 presented results in reading aloud. For experienced interpreters, PBD units required comparatively less time in the first pass but more time in the second pass and beyond. Nonetheless, dependent *t*-tests confirmed that non-PBD and PBD units took statistically similar amount of time across all stages of processing. Interpreting students, except FFD, had consistently longer duration on PBD units regarding all indices. Still, statistical analyses came back with a non-significant effect for all indices. Untrained bilinguals spent less time on PBD units when FFD and GPT are considered, while GD, RRT, and TVT on the same type of unit were prolonged. However, no significant effect was found for this group in reading aloud, either.

Table 25 Mean fixation duration on non-PBD and PBD units in reading aloud across groups (unit: millisecond)

	PRO		NEW		BIL	
	Non-PBD	PBD	Non-PBD	PBD	Non-PBD	PBD
FFD	286.896	273.298	296.010	288.889	291.995	274.536
	(32.761)	(23.143)	(33.729)	(43.534)	(29.470)	(30.205)
GD	350.759	343.865	362.600	375.457	352.331	357.441
	(24.537)	(25.532)	(26.648)	(51.094)	(30.452)	(47.352)
GPT	412.047	419.741	464.406	478.238	467.722	448.852
	(50.869)	(43.820)	(79.537)	(120.224)	(69.971)	(80.679)
RRT	276.643	307.497	360.001	360.020	331.209	366.941
	(83.303)	(100.524)	(93.819)	(185.075)	(71.784)	(86.772)
TVT	397.734	407.636	454.930	487.366	452.918	463.642
	(40.355)	(41.212)	(67.675)	(147.055)	(71.717)	(60.291)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*FFD = First fixation duration; GD = Gaze duration; GPT = Go-past time; RRT = Rereading time; TVT = Total viewing time
*PBD = Principal branching direction

Lastly, the focus lands on sight translation, in which the effects of contrastive syntactic structures were suspected to be even larger, due to the necessity of reformulation. Table 26 provides the general results in sight translation. Experienced interpreters and interpreting students differed only in that GPT on non-PBD units for the latter group was slightly longer. This aside, PBD units appeared to have consistently taken longer time to process for both groups. Yet again, statistical analyses demonstrated that the two types of units took just similar amount of time.

When it comes to untrained bilinguals, the numerical pattern also indicated a longer time for all indices on PBD units, but statistical analyses showed diverse results from the other two groups. For one, GD on non-PBD units (M = 264.086) was significantly shorter than that on the PBD counterparts (M = 304.959), $t(6) = -4.876$, $p = .003$. In addition, RRT also turned out to be significantly shorter on baseline units (M = 537.316) than on words embedded in PBD structures (M = 625.67), $t(6) = -2.75$, $p = .033$. The same applied to TVT: The mean duration was also significantly shorter on non-PBD units (M = 623.06) than on PBD units (M = 722.066), $t(6) = -3.455$, $p = .014$.

Table 26 Mean fixation duration on non-PBD and PBD units in sight translation across groups (unit: millisecond)

	PRO		NEW		BIL	
	Non-PBD	PBD	Non-PBD	PBD	Non-PBD	PBD
FFD	232.477	239.794	249.351	255.661	238.551	249.081
	(36.039)	(23.379)	(32.926)	(26.023)	(37.705)	(44.213)
GD	239.570	251.460	273.942	298.340	264.086	304.959
	(37.767)	(26.068)	(44.706)	(34.939)	(40.662)	(46.658)
GPT	286.208	323.041	464.237	458.769	331.795	368.489
	(76.909)	(64.186)	(219.904)	(160.182)	(82.621)	(91.808)
RRT	326.964	402.927	505.696	527.488	537.316	625.670
	(142.299)	(127.233)	(190.731)	(215.880)	(164.574)	(202.760)
TVT	412.183	494.787	569.138	605.781	623.060	722.066
	(163.935)	(142.475)	(235.374)	(239.702)	(216.771)	(246.882)

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*FFD = First fixation duration; GD = Gaze duration; GPT = Go-past time; RRT = Rereading time; TVT = Total viewing time
*PBD = Principal branching direction

From the numerical patterns in all three kinds of tasks, we have witnessed that divergent PBD structures in English and Chinese demanded longer time to process, which was true for all groups. The finding had been readily expected, since in English these PBD structures (i.e. adjuncts) follow the antecedents and hence are in direct contrast to the way elements are organized in Chinese. In this way, extra effort is needed to hold the information about the antecedent and link it to its adjunct for native speakers of Chinese. That is the reason why we saw a significantly longer time on PBD units in the stage of meaning interpretation and integration for untrained bilinguals in silent reading.

Moving on to sight translation, the gap between the processing loads of non-PBD units against PBD units widened, as was speculated. This time, in addition to GD and RRT, TVT on PBD units was also significantly longer for untrained bilinguals. The need to speak out the result of reformulation means that a general understanding would no longer suffice, and therefore forced the untrained participants to strive for an explicit, accurate understanding of the links between antecedents and adjuncts and then come up with appropriate correspondents in another language. Accordingly, PBD units were viewed specifically for more times so that they could be translated and inserted into suitable positions in the process of output production.

On the flip side, experienced interpreters and interpreting students not only successfully narrowed the gap between the processing of non-PBD and PBD units in silent reading, but also managed to keep the gaps non-significant all the way through the task of sight translation.

4.5 Quality revisited: factors that had a say

Previously we reported that the quality differences between every two groups were all statistically significant. While this is true for accuracy, for fluency, the other sub-indicator that decided the overall quality, interpreting students were actually on par with experienced interpreters. The sole divide between groups regarding fluency hinged on training, so the bilinguals were the only group falling behind. Factors that might have some influence on quality — and further, accuracy and fluency — were subject to a multiple regression analysis, and the results are presented below respectively.

Table 27 summarizes all the factors subject to statistical analysis and their impacts on the overall quality of output, including “group”, “total time”, “reading span”, “word counts”, “reading-ahead span (1st utterance)”, “reading-ahead span (non-1st utterance)”, “length of verbal gaps”, and “percentage of pause”. Overall, the analysis found a significant result, $R^2 = .834$, $F(8,44) = 33.699$, $MSE < 1$, $p < .001$. It turned out that only the factor of “group” would significantly sway the outcome of sight translation output, $p < .001$, accounting for the biggest impact (*beta* = -.353). The factor was negatively related to quality, denoting that experts scored the highest (Group code = 1) and bilinguals scored the lowest (Group code = 3). Collinearity diagnostics found some collinear relationship between total time (.87), word counts (.85), reading-ahead span before the 1^{st} utterance (.69), and also non-first-utterance reading-ahead span (.60).

Table 27 Factors that have a significant effect on overall quality

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	SE	Beta	T	Sig.	Tolerance	VIF
(Constant)	6.294	.704		8.939	.000		
Group	-.353	.091	-.353	-3.875	.000	.384	2.602
Total time	-.003	.003	-.321	-.951	.347	.028	35.858
Reading span	.062	.056	.069	1.095	.280	.806	1.241
Word counts	-.002	.001	-.156	-1.633	.110	.352	2.844
Reading-ahead span (1st utterance)	-.006	.003	-.250	-1.613	.114	.133	7.534
Reading-ahead span (non-1^{st} utterance)	.057	.268	.059	.214	.832	.042	23.863
Length of verbal gaps	-.002	.002	-.223	-1.273	.210	.104	9.619
Percentage of pause	.001	.021	.006	.045	.964	.155	6.468

Results for the multiple regression analysis of accuracy are provided in Table 28. A main effect was significant, R^2 = .665, F(8,44) = 13.89, *MSE* < 1, p < .001. The explanatory power wasn't really high, and only around 66.5% of the data could be explained by these variables. This is readily expected, since many indices were more or less related to one's speed and fluency. Group was again found to be the only significant factor that impacted the outcome of accuracy, p < .001. This time, the effect was even more potent than it was on the overall quality (*beta* = -.514).

Table 28 Factors that have a significant effect on accuracy

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	SE	Beta	T	Sig.	Tolerance	VIF
(Constant)	6.422	.977		6.576	.000		
Group	-.501	.126	-.514	-3.966	.000	.384	2.602
Total time	-.001	.004	-.194	-.403	.689	.028	35.858
Reading span	.094	.078	.107	1.196	.238	.806	1.241
Word counts	-.002	.002	-.150	-1.107	.274	.352	2.844
Reading-ahead span (1st utterance)	.000	.005	.018	.084	.934	.133	7.534
Reading-ahead span (non-1^{st} utterance)	-.075	.372	-.079	-.201	.842	.042	23.863
Length of verbal gaps	-.004	.002	-.439	-1.761	.085	.104	9.619
Percentage of pauses	.047	.029	.332	1.625	.111	.155	6.468

Results for the multiple regression analysis of fluency are presented in Table 29. An overall significant effect did exist, R^2 = .877, F(8,44) = 47.415, MSE < 1, p < .001, and three factors were found to be at work, including "group", p = .027, "reading-ahead span (1st utterance)", p = .001, and "percentage of pauses", p = .033. All three factors were negatively related to the ultimate fluency score. That is, the fewer pauses detected and fewer fixations used before starting to sight-translate the text, the higher the fluency. What's more, the reading-ahead span before oral rendition begins was the most decisive variable.

Table 29 Factors that have a significant effect on fluency

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	SE	Beta	T	Sig.	Tolerance	VIF
(Constant)	6.166	.690		8.937	.000		
Group	-.205	.089	-.180	-2.295	.027	.384	2.602
Total time	-.004	.003	-.399	-1.369	.178	.028	35.858
Reading span	.030	.055	.029	.541	.591	.806	1.241
Word counts	-.002	.001	-.145	-1.765	.084	.352	2.844
Reading-ahead span (1st utterance)	-.011	.003	-.455	-3.411	.001	.133	7.534
Reading-ahead span (non-1st utterance)	.189	.263	.171	.720	.475	.042	23.863
Length of verbal gaps	.000	.002	-.016	-.105	.917	.104	9.619
Percentage of pauses	-.045	.020	-.273	-2.208	.033	.155	6.468

The factor of group continued to play an important role in all three multiple regression analyses. For the time being, accuracy can only be explained by group differences, but more factors can serve as our reference in terms of fluency.

Looking closer at the indices used in this study, we found that bilinguals always stayed at one extreme in terms of almost all indices in sight translation, except word counts. What's really astonishing is what happened to interpreting students. Though "bottoming out" in silent reading and reading aloud, trainees actually closed in on experienced interpreters and left bilinguals far behind in sight translation, reflected through a comparable fluency with experts and a significant better accuracy and fluency than the untrained participants.

4.6 Summary

This chapter reported results of global observations, local reading indices, and behaviors during the process of sight translation in sequence. As a result, three groups were generally similar in silent reading and reading ahead with equal level of comprehension, except that experienced interpreters often stood out with a better performance (i.e. faster processing speed). The time when the impacts of training and experience really showed themselves was in both the process and the outcome of sight translation.

From a macro perspective, trained participants required fewer fixations and less time to finish the task of sight translation, with a superior quality, albeit all three groups used similar amount of words. Training gave interpreting students a competitive edge — reflected through surpassing accuracy and fluency — over untrained bilinguals. Experience, on the other hand, left a trace of its influence in that experienced interpreters scored higher accuracy than trainees.

From a micro perspective, there wasn't any divergence in the local reading indices. That is, word-based processing efforts were similar in sight translation between all three groups. The real differences lay in the behavior in the process, and we saw clear evidence of the impacts of training and experience in successive stages. Starting from the moment the text appeared before the participants till the task of sight translation was over, the influence of training was dominant. While most untrained bilinguals read through the text and remained silent for a long time before oral rendition began, most trained participants only had few looks and started sight-translating the text almost immediately.

Along the process, trained participants consistently had fewer and shorter observable pauses and faster output speed, and they needed fewer fixations to acquire information during each verbal gap and pause as well to make it through the task. One thing worthy of notice is that the effects of experience kicked in once the production phase began. After an comparable start with the interpreting students, experienced interpreters outperformed trainees in that their pauses were even fewer and shorter, and their output was even faster. On top of that, experts read even fewer times than trainees while simultaneously sight-translating the text and rarely seemed slowed down by difficulties they encountered.

Lastly, regardless of the background of the participants, adjuncts with opposite principal branching directions (PBD) in English and Chinese attracted much attention of all people during pauses, though trained groups did not differ in the efforts spent on PBD units against non-PBD units as the untrained group did.

Chapter 5 Discussion

This chapter is divided into several sections to address each of the research questions set in this study. Firstly, how reading purposes change reading behavior will be discussed. Next, we focus on how training weighs in and further affects reading and sight translation behavior. After elucidating that training is efficient and effective in behavior changing, we turn to experience and unveil how sight translation is brought to a different level, offering some implications for interpreter training. Features in the process of sight translation and even pauses are put in the last two sections as supplementary information.

Reading span (working memory) has been excluded in this chapter, since statistical analysis already confirmed that working memory neither differed significantly between groups nor contributed to the variations in accuracy and fluency scores. Logical relationships between tasks for each group are provided in Table 30 as a reference when we discuss between-condition comparisons in Section 5.1 below.

Table 30 Between-condition comparisons across groups[5] (reference table)

	PRO	NEW	BIL
FFD	SR=ST<RA	ST=SR<RA	ST=SR; SR<RA ST=RA
GD	ST=SR<RA	ST=SR<RA	ST=SR<RA
GPT	SR<RA SR=ST; ST=RA	SR=ST= RA	SR=ST=RA
RRT	SR<RA=ST	SR=RA=ST	SR=RA<ST
TVT	SR<RA=ST	SR=RA=ST	SR=RA; RA=ST SR<ST
Rereading count	RA=SR; SR=ST RA<ST	RA=SR; SR=ST RA<ST	RA=SR<ST
No. of fixations	SR=RA<ST	SR=RA<ST	SR=RA<ST
Total time	SR=RA<ST	SR=RA<ST	SR=RA<ST

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*FFD = First fixation duration; GD = Gaze duration; GPT = Go-past time; RRT = Rereading time; TVT = Total viewing time
*SR = Silent reading; RA = Reading aloud; ST = Sight translation

5.1 Reading for comprehension vs. reading for sight translation: The impact of reading purpose

The similarities between the three groups are visualized in Figure 4 below. During first-pass reading, both FFD and GD proved that, for all three groups, silent reading was similar, and almost identical, to sight translation, while both were significantly shorter than reading aloud. The finding suggests that the additional task of vocalization required in reading aloud had already taken its toll on all the participants in an early stage, corroborating Huang’s (2011) findings (see also Rayner, 2009).

[5] Numerically speaking, the task placed on the left of any logical symbol is always lower than the one on the right. In addition, statistical significance is represented with “>” or “<”, and “=” stands for non-significance.

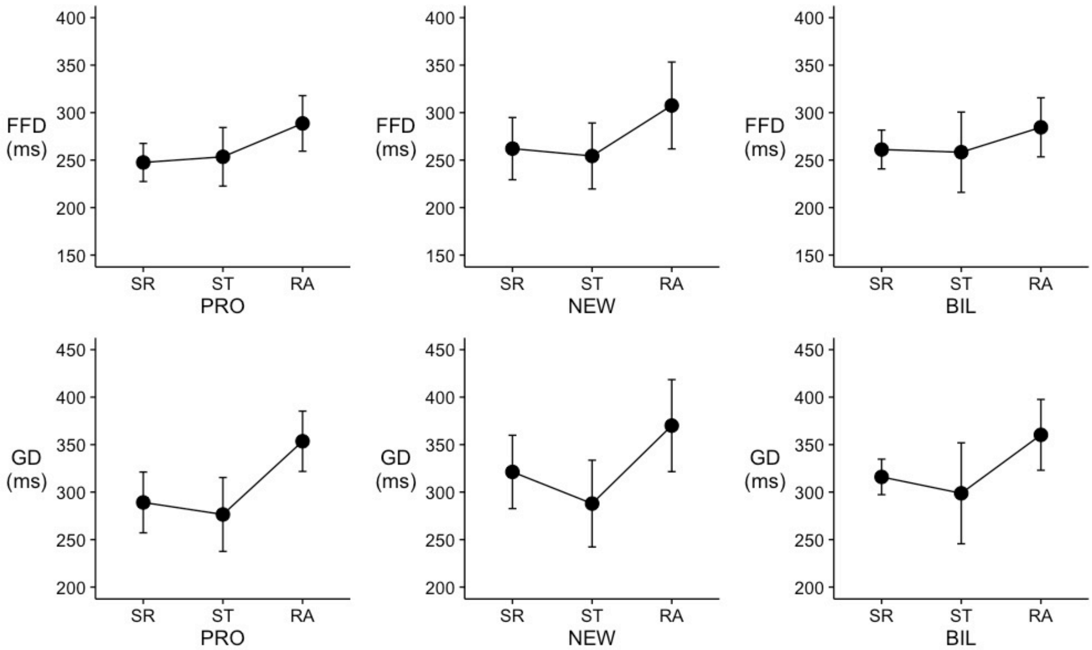


Figure 4 First-pass reading indices across conditions for all groups

*The error bar in figure 5 marks the standard deviation for each group

However, Chen's (2013) research on experienced interpreters found first-pass indices in sight translation to be significantly longer than those for silent reading, and thus speculated that her participants were actually reformulating messages during the first pass. In this study, no significance was found. The most plausible answer to this inconsistency may be the working direction of sight translation, in which our studies happen to be opposite. Chen (2013) studied Chinese-to-English sight translation. With the source text in one's native language, comprehension might be short enough that, when additional tasks were performed, extra time spent could easily lead to a significant difference. On the contrary, participants in the current study were required to read in their foreign language. With longer time needed for comprehension and accelerated retrieval of equivalents in one's first language (cf. Frenck-Mestre, 1997, 2002), the gap might not be large enough to create a significant difference between silent reading and sight translation in the first pass. This line of extrapolation and Chen's assertion actually both have their support from our data, though not entirely conclusive. In this study, experienced interpreters were quite distinct in their strategy of tackling sight translation, as far as FFD is concerned. While trainees and bilinguals chose to limit their time spent on reading new information in sight translation (reflected in shortened FFD), experts, on the opposite, devoted slightly more time, leading to a longer FFD, when compared with silent reading. It therefore seems likely that, as Chen proposed, experts were engaging in additional work during this earliest stage of processing. However, we have no way to know whether the additional processing entails reformulation or just part of it with the data currently available.

The incongruity between Chen's (2013) finding and ours could also come from the disparity between participants because all our experts went through formal interpreting training, unlike some of Chen's experienced participants, although work experience as a threshold for selection were both set at 150 days of work. Still another possibility is that the experts in this study did not embark on additional processing other than simply reading and left reformulation and production in the second pass and onwards (as described by McDonald and Carpenter, 1981). Or, even reformulation began in the first pass, the final decision of reformulation and the execution of oral production lingered on to later stages.

An interesting observation is that the trainees' behavior here did not corroborate Huang's (2011) finding, in which trainees echoed experts' approach. Again, we would like to propose that this may be the consequence of an opposite working direction. Reading in one's native language gave participants in Huang's study some confidence to begin next-stage processing at the same time when they were trying to comprehend the message, while in this study, a less familiar language compelled trainees (and also bilinguals) to adopt a safer technique to make sure they get the message during the first encounter.

It's also worth noting that, for bilinguals, though the duration of FFD in both silent reading and sight translation were almost identical. The wider standard deviation in sight translation, along with the reading behavior before beginning interpreting, may to some extent support Shreve, Schäffner, Danks, and Griffin's (1993) assertion that readers tend to read more thoroughly and deliberately when tasks in addition to comprehension are required, but in this study only participants without training fit the profile.

In terms of GD, all three groups this time exhibited the same pattern. Comparing silent reading and sight translation, we found that GD decreased across all groups, though only numerically. This phenomenon is again different from Huang's (2011) and Chen's (2013) findings. In their Chinese-to-English sight translation studies, GD was longer in sight translation, though only experts' GD reached a significant difference between tasks. This brings us to one question: If experts did resort to a different approach, why did GD decline, as in the other two groups?

There are two possibilities. The first is that experts were indeed acting just like others, leaving all reformulation and production work for the second pass and beyond. Therefore, FFD was longer, but GD revealed their true intent, which was to speed up so they would not take too long before coming back for reformulation and production. There is also another way to look at the finding, that is, the longer FFD did reflect some additional processing. As for the shorter GD, the data from between-group comparisons should be taken into consideration. In silent reading, experts' GD was significantly shorter than that of trainees' and bilinguals'. Nonetheless, in sight translation the gap was narrowed to a non-significant level. This change obviously entailed additional work of some sort for experts. But still, an awareness of the fact that there was audience waiting might still explain a shorter GD in sight translation than in silent reading for experts. On the other hand, the largely decreased GD for trainees and bilinguals are considered only as the consequence of being aware of the audience. In order to set aside some time for reformulation without keeping the audience waiting for too long, these two groups chose to largely reduce the duration of GD, so they could begin their second reading pass earlier.

As illustrated in Figure 5, from GPT on, sight translation's burden started to become visible, that is, GPT in this task was no longer the shortest for any group, and it even had the widest standard deviation across the three tasks. For experts, silent reading was equal to sight translation. As for trainees and bilinguals, three tasks were deemed equal by statistical analysis, but reading for comprehension and reading for sight translation started to go separate ways.

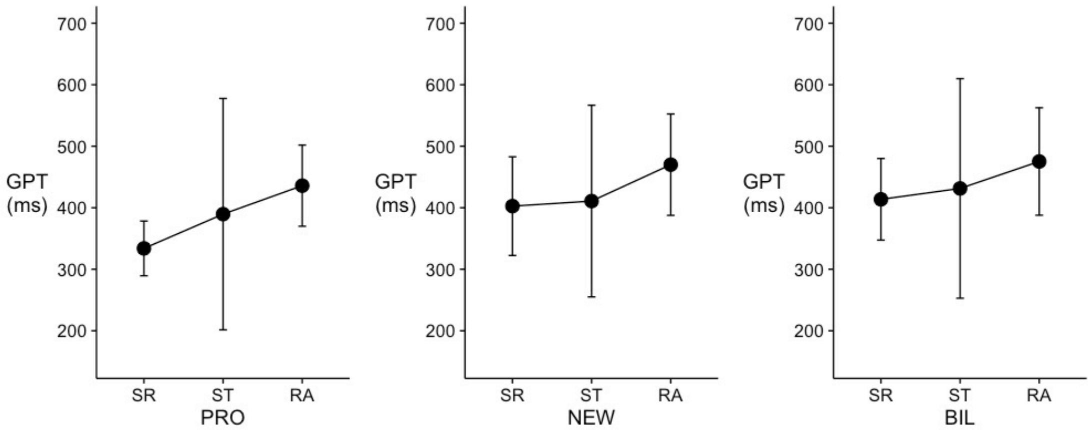


Figure 5 GPT across conditions for all groups

*The error bar in figure 5 marks the standard deviation for each group

Going further into other non-first-pass indices, we found more contrasting features, as pointed out more clearly in Figure 6. For experienced interpreters and untrained bilinguals, RRT in silent reading and in sight translation began to diverge significantly. The need to go back to previous regions for reformulation and production in sight translation is clearly reflected through a longer RRT. This may point to the fact that, even for experts, sight translation in the first reading pass does not seem to be the norm, but only an additional advantage brought by experience sometimes.

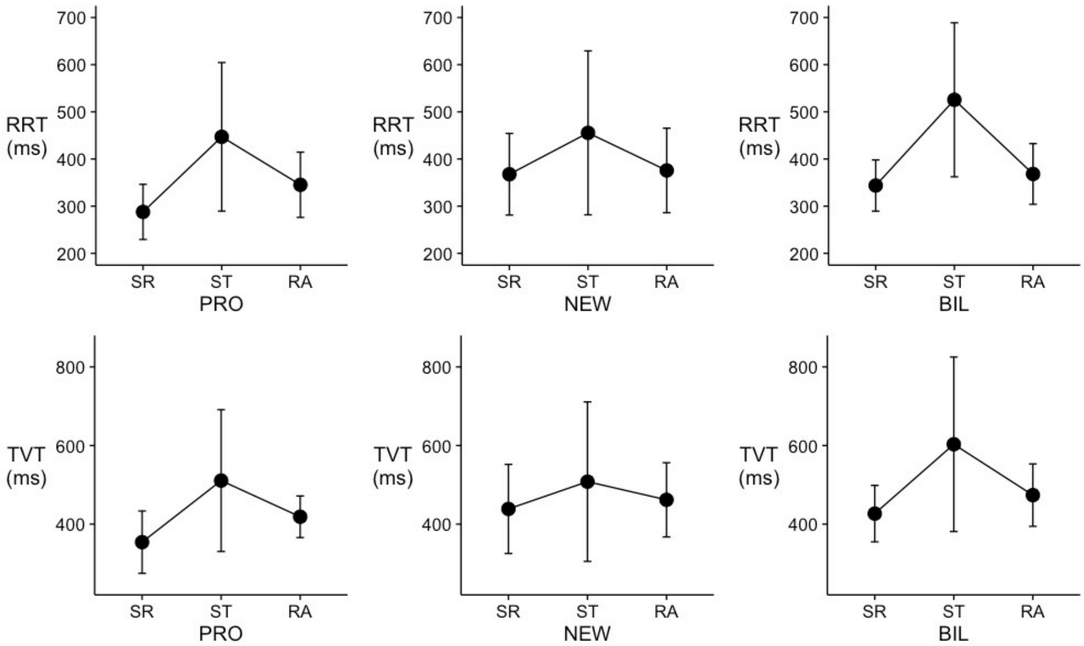


Figure 6 Non-first-pass indices across conditions for all groups
*The error bar in figure 6 marks the standard deviation for each group

All three tasks were statistically similar for interpreting students, though sight translation was still the most time-consuming task and had the widest distribution. The fact that this group of participants almost consistently had the longest duration in most of the reading indices in silent reading and reading aloud is perhaps the reason why none of the gaps reached significance. This finding contradicts Huang's (2011) study, in which interpreting students had significantly shorter RRT in silent reading than in sight translation. This again may be the effect of a different language direction.

The TVT of experienced interpreters and of bilinguals both demonstrated that reading for sight translation is an outright different task from reading for comprehension. On the other hand, interpreting students still showed no difference between all three tasks. There might be two possible reasons. The first is that the shorter FFD and GD in sight translation than in silent reading occurred as a function of a deliberate action, hence the similarity between tasks. The other possibility is that trainees had a rather long TVT in silent reading, and the TVT in sight translation was not long enough for the gap to reach a significant difference.

Generally speaking, due to the opposite working direction of sight translation, our findings here vary greatly from those of Huang's (2011) and Chen's (2013). The main reason behind may be the efficiency of information retrieval (cf. Frenck-Mestre, 1997, 2002). With one's first language as the source content, effects of various tasks can be easily seen. However, in the context of foreign language reading, even a slightly more amount of effort devoted towards word decoding might obscure any effect that might be at work in different tasks. Some may suggest that the inconsistency of the difficulty level between the materials used in this experiment and those in Huang's (2011) and Chen's (2013) theses might account for the varying research findings. The materials were more formal in this experiment but more colloquial in the other two studies. Indeed, the difficulty level of the texts might induce obscuring effects. However, experienced interpreters, among all groups, could not be more familiar with the register and the domain knowledge of the context chosen for this study, due to their interpreting expertise. Hence, we are still convinced that the opposite language direction might better explain the disparity in the experts' divergent behaviors in different studies.

Looking at a bigger picture, the total number of fixations and total time reflected even more obvious disparities. While silent reading was similar to reading aloud, the two tasks were each significantly fewer and shorter than sight translation in the above two measures for all three groups.

Previous studies, such as McDonald and Carpenter (1981) and Shreve, Schäffner, Danks, and Griffin (1993), contended that reading for comprehension is in general similar to reading for sight translation. With regard to our statistics, these two tasks did not differ significantly in FFD and GD, so we can safely assume that for bilinguals (and maybe interpreting students), reading purposes did not affect the features of first-pass reading, and there was only an acceleration of speed in sight translation. However, this cannot be said to be definitive for experienced interpreters because they were clearly putting in additional effort in things other than reading for comprehension.

Does that render scholars that hold the opposite opinion wrong? In fact, our global indices and also non-first-pass indices support studies that have claimed reading for comprehension to be different from reading for sight translation, or that reading purposes do impact reading behavior (see Alves, Pagano & da Silva, 2011; Dragsted and Hansen, 2009; Göpferich, Jakobsen, and Mees, 2008; Macizo and Bajo, 2004; Macizo and Bajo, 2006; and Shreve, Lacruz & Angelone, 2010).

Accordingly, based on the findings in this study, we may suggest that reading purposes do affect reading behavior. While these two kinds of reading are similar in the first pass, they start to differ in the second pass and beyond, with the burden of reformulation and oral production borne by sight translation clearly reflected in fixation duration. On the other hand, reading aloud makes its presence felt by putting additional burden early in the process, which can be identified through first-pass reading measures.

That said, ultimately experience may change one's reading behavior, just like what our experts did at the first encounter with words. Training, in the midst of this discussion, may also play an important role. In addition to a similar RRT in all tasks for interpreting students, rereading counts of both trained groups of participants point out an even more remarkable feat of training. Bilinguals' data perfectly fit our analysis that non-first-pass measures differentiate silent reading from sight translation, but trainees and experts successfully brought down rereading counts in sight translation to the same level as silent reading. In other words, the task of reformulation and production did not require significantly more rereading, which either means trained participants were more competent or decisive in finding an appropriate rendition after the first pass, or that the source text was sometimes sight-translated at the first encounter, deviating from the McDonald and Carpenter (1981) model. As far as we are concerned, both explanations seem equally possible. Take interpreting students as an example. The group of trainees had shorter FFD and GD to, in our opinion, speed up reading in order to come back for reformulation as soon as possible. If this strategy had been carried out strictly, rereading counts in sight translation should have been significantly more than silent reading, since reformulation would double passes of reading each time. Yet, findings found that trainees had similar rereading counts in both silent reading and sight translation. Therefore, we would suggest that the real situation is likely a combination of both.

Still, we can only offer the above speculation, arguing that a possible difference in behavior may be at work behind the scenes because we do not have enough data for now. To really find out whether experience and training would change reading behavior, we need to map out the eye movements in silent reading and sight translation and compare how and where they move in the two tasks.

5.2 The influence of training on output and reading behavior

Knowing that reading purposes do change reading behavior, we now turn our focus to how training exerts its impacts. Table 31 compiles results entailing all tasks and simplifies the raw data, presenting them in logical relationships instead. Discussion in this section centers on the influence of training, and hence the attention will be given mostly to interpreting students and bilinguals.

Table 31 Between-group comparisons across tasks[6] (reference table)

	Silent reading	Reading aloud	Sight translation
FFD	PRO=BIL=NEW	BIL=PRO=NEW	PRO=NEW=BIL
GD	PRO<BIL=NEW	PRO=BIL=NEW	PRO=NEW=BIL
GPT	PRO<NEW=BIL	PRO=NEW=BIL	PRO=NEW=BIL*
RRT	PRO<BIL=NEW	PRO=BIL=NEW	PRO=NEW=BIL*
TVT	PRO<NEW BIL=NEW PRO=BIL	PRO=NEW=BIL	NEW=PRO=BIL*
Rereading count	PRO=BIL=NEW	PRO=NEW=BIL	PRO=NEW=BIL
No. of fixations	PRO=NEW NEW=BIL PRO<BIL	PRO=NEW NEW=BIL PRO<BIL	PRO=NEW<BIL
Total time	PRO=NEW NEW=BIL PRO<BIL	PRO=NEW NEW=BIL PRO<BIL	PRO=NEW<BIL

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*FFD = First fixation duration; GD = Gaze duration; GPT = Go-past time; RRT = Rereading time; TVT = Total viewing time

Globally, interpreting students and bilinguals had similar number of fixations and total time during silent reading and reading aloud. Furthermore, all the reading indices (FFD, GD, GPT, RRT, and TVT) turned out to be similar, and trainees in fact spent numerically longer time on all five reading indices.

The aforesaid findings can prove that the two groups are identical in their general linguistic competence, and the difference between two groups lies only in their method of tackling texts: Trainees spent longer time locally, which were then compensated with shorter task time and fewer fixations globally, while bilinguals had

[6]Numerically speaking, the task placed on the left of any logical symbol is always lower than the one on the right. In addition, statistical significance is represented with “>” or “<”, and “=” stands for non-significance.

a reverse pattern. Only the participants that passed the threshold of comprehension test were selected was another fail-safe mechanism to ensure same level of understanding of the texts. Hence, we can say with some confidence that significant differences in indices for sight translation may come from training. As was expected, global indices such as total number of fixations and total time in sight translation showed that interpreting students were more efficient when tackling the task of sight translation, and the accuracy and fluency ratings of the output were also significantly better than bilinguals, further attesting to the effect of training (see Table 32 below).

What led to a shorter total time and fewer total fixations? Is it possible that a more concise style of rendition was responsible? It turns out that this explanation doesn't add up because total word counts of participants' output show that there was no significant difference across groups. What about the efficiency of information extraction? Did reading indices for sight translation exhibit a different pattern from the other two tasks? As a matter of fact, this study found that all reading indices, including FFD, GD, GPT, RRT, and TVT all failed to show significant differences between interpreting students and bilinguals. Therefore, efficiency of meaning extraction seems to be similar between the two groups. What stands out here is the different patterns in sight translation. Trainees consistently had non-significant, shorter local indices and even significantly shorter global indices. It's obviously the effect of training that gave trainees an advantage, securing an overall win over bilinguals in speed.

Another interesting fact is that, while first-pass indices remained close for trainees and bilinguals, the gap gradually widened along different stages of reading, at last reaching the pinnacle in TVT. During training, interpreting students time and again practiced the skill of chunking to “translate along the text” so as to shorten the time needed to go back to previous regions. Thus, much shorter later-stage indices may hint at the possible influence of training because trainees learned to try to engage in reformulation as soon as possible. This strategy can also adequately explain why trainees had fewer fixations in total.

Aside from fewer fixations, we may turn to other reading behavior in the process of sight translation for possible contributors to a less total time. Indices specifically targeting sight translation are listed in the following Table 32.

Table 32 Sight translation indices and reading behavior across groups[7] (reference table)

Sight translation indices & Reading behavior in the process	
Quality	BIL<NEW<PRO
Quality: accuracy	BIL<NEW<PRO
Quality: fluency	BIL<NEW=PRO
Word count	NEW=PRO=BIL
Reading-ahead span (1st utterance)	NEW=PRO<BIL
Initial silence	NEW=PRO<BIL
Reading-ahead span (non-1st utterance)	PRO<NEW<BIL
Percentage of pause	PRO<NEW<BIL
Pause length	PRO<NEW<BIL
No. of fixations (by pause)	PRO<NEW<BIL
Verbal gap duration	PRO<NEW<BIL

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals

It was found that the initial silent period and number of fixations before participants began interpreting in each trial were both significantly less for interpreting students. On top of that, the distribution was incredibly dense for the group, meaning that almost all trainees exhibited the same tendency to avoid keeping the audience waiting. Bilinguals, though might also be aware that there was audience waiting, clearly chose an opposite strategy: Most read through the trial before opening their mouth. This might be that, without the knowledge of chunking, bilinguals had good enough linguistic capability to know that they would be better off getting all the information in advance.

[7] Numerically speaking, the task placed on the left of any logical symbol is always lower than the one on the right. In addition, statistical significance is represented with “>” or “<”, and “=” stands for non-significance.

Once participants began sight-translating the text, training again showed its impact. Trainees on average read significantly less than bilinguals during each non-first utterance. This finding is reasonably expected, since the strategy of "translating along the text" gave trainees the ability to translate with minimal meaningful units at hand, thus largely reducing the need to read too far ahead before coming back to the same text again for reformulation. As we have speculated, some reformulation and production might have been completed in the first pass of reading because rereading counts in sight translation was similar to the number in silent reading for trainees. Thus, by far our findings seem to corroborate McDonald and Carpenter's (1981) three-stage model, as did Su (2013), but with some exceptions.

Of all the verbal gaps in each trial, 13.9% were sensed as pauses for interpreting students, significantly less than bilinguals' 19.44%. In addition, average pause length for trainees was significantly shorter. Cross-referencing the average number of fixations each group had during each pause, we found that trainees differed significantly from bilinguals in 25% of the pauses because both groups had only one fixation in 75% of the pauses. Taking the average pause length and number of fixations together, trainees generally reacted more quickly and required fewer fixations to help them continue. Still, it seems that the trainees were still getting used to the skill of chunking, and therefore had to slow down sometimes to combat problems encountered, as the bilinguals did. This extrapolation can at least be partly substantiated by the max values for the two groups on average pause length and the number of fixations in each pause.

Lastly, working pace of sight translation reflected through the average duration of verbal gaps showed that interpreting students was significantly quicker than bilinguals. Interestingly, most bilinguals read through each trial before they began interpreting, while trainees mostly started with just a few peeks, but along the way trainees could still resume interpreting quicker than untrained bilinguals after they paused, which is another clear evidence of the impact of training.

From the statistics, we saw that the reading pattern and the time saved along the process helped lead to fewer fixations and less task time in total, and fewer fixations here also means fewer reading passes, since average duration of each reading index was statistically similar. What is really intriguing is that bilinguals and trainees might have different mechanisms. For most of the bilinguals, first-pass reading was much more about comprehension, as they tended to read through the text first; on the contrary, the same pass of reading might have entailed both comprehension and reformulation (or at least part of it) for trainees.

5.3 The influence of experience on output and reading behavior

If training has immediate effects on both the process and outcome of sight translation, does experience has anything more to offer? What does the next level of sight translation look like? In this section, we turn to comparisons between experienced interpreters and interpreting students to outline the impact of experience.

A look at measures in silent reading for an identification of linguistic competence in the first place is in order. Global indices including the total number of fixations and total time were similar for both groups. However, every local reading index showed a significant difference, except for FFD. This is reasonable evidence that experienced interpreters were more efficient in meaning retrieval, and hence a probable superiority in general linguistic abilities. In contrast, Chen (2013) found a longer duration in GD, RRT, and TVT for trainees in silent reading, but none of the gaps in the task reached significance when compared with experts. This again tells us that language direction is an important factor to reckon with when making interpretation of the results.

Turning to sight translation, however, all global and reading indices failed to show a difference between the two groups. It could be that training considerably boosted interpreting students' speed, so the disparities apparent in silent reading were no longer observable, or the results were simply masked by the additional burden of reformulation and production conducted in different stages by each group.

We would like to direct our discussion to the fluctuations of numerical differences. In early reading indices, the divide between trainees and experts broadened gradually, reaching a maximum in GPT. From this we can infer that trainees adhered more strictly to the basic chunking routine, shortening the first pass and immediately going back to the initial positions of meaningful units for reformulation (see previous section in which we discussed the acceleration of trainees in the first pass). In the meantime, experts' shorter GPT may indicate the fact that a higher percentage of the message had been tackled in the first pass without regression, as proclaimed by Hsieh (2014), revealing the trace of experience.

However, going into later stages, trainees and experts started to behave much alike in terms of RRT and even TVT. At this point, bilinguals' much longer duration may serve as a great baseline for comparison. Trained participants were more confident and decisive in that they moved along the text successively, segment by segment, without frequently going back to previous regions once they finished sight-translating the meaningful unit in their central vision, while untrained bilinguals were going back and forth continuously to piece together information in a correct order. The above deduction can also be substantiated by the rereading counts in different tasks for each group (see in Table 28 that both experts and trainees were able to limit rereading counts in sight translation to the same level as silent reading, but not bilinguals). In this regard, training seems to help develop a segment-by-segment approach, while experience is more visible only in the relative efficiency and effectiveness of handling each segment.

How about the reading pattern during sight translation? Can we see the impact of experience on what happened along the way? Starting from the number of fixations and length of silence before the participants began to deliver for the first time, we found that the two groups were statistically similar, except that professionals showed more variations in style. Trainees were extremely homogeneous, while experience seemed to free some experts to a certain extent from this "don't-keep-the-audience-waiting doctrine."

The role of experience became more tangible once interpreting began. Experienced interpreters on average needed significantly fewer fixations ahead of each non-first utterance, and the distribution was much narrower, gravitating around 1.5 fixations, indicating that the style was consistent across all experts. The percentage of pause among all verbal gaps was significantly lower for experts (9.772%) than trainees (13.9%), exhibiting a smoother sight translation output, which could be further supported by a significantly shorter length of pauses on average. Even if experts indeed ran into problems, they didn't dwell on the issues for too long. After a pause of 1.474 seconds at most on average, experts kept the ball rolling again. On the other hand, trainees would need to stop for some time before moving on when real problems turned up. The number of fixations during each pause confirms this observation with a significant gap between the two groups. Experience perhaps gives professionals more flexibility in their choices for reformulation, so they could almost always quickly maneuver around the obstacles.

With much fewer and shorter pauses, experienced interpreters were still able to maintain a faster pace when sight-translating the text. Average length of verbal gaps showed a significant shorter length for experts than interpreting students. Thus, we can confidently say that professionals were generally faster, not just in situations that required problem solving. The flexibility honed by years of practice enabled a smoother and quicker rendition when experts translated along the text.

From all the patterns manifested during sight translation, we can sense that trainees were trying to rely on the same chunking strategy as experts did, and they actually mastered the "form" (the method of chunking), so the only difference lay in the “content”, that is, the flexibility in language, acquired through experience, that seamlessly combine chunked messages together in a coherent way.

So far, the influence of training is much more visible than that of experience because, with training, interpreting students did successfully catch up with experienced interpreters in many aspects, including almost all global indices, local reading indices, and even rereading counts. Our findings coincide with those of Chmiel and Mazur’s (2013), in which trainees, regardless of the length of interpreter training, finished the sight translation task with comparable numbers of fixations, mean fixation duration and total time. This convergence of evidence proves that a formal training in sight translation skills is enough to wield an immediate, positive influence.

However, there are still some insurmountable gaps. First and foremost, though fluency was similar between experts and trainees, accuracy still clearly portrayed the value of experience, just as McDonald and Carpenter (1981) found in their study that experience led to a boost in accuracy, not speed. Lee (2012) also provided corresponding evidence. In Lee’s experiment, experts had fewer major errors and expression errors, both of which go under the category of accuracy in the current study. In addition, experts also spent less time on the task. It’s a pity that no statistical method was used to verify the numerical differences in Lee’s study, or we might have more solid support for our findings.

Aside from higher accuracy as a group, the real difference appeared after participants began interpreting. Experts were able to maintain a faster pace of interpreting, and they stuttered or hesitated fewer times. Even when experts stumbled, they paused for much less time and needed fewer fixations to get out of trouble. What's more, experts on average relied on largely fewer fixations between every two utterances to keep things going smoothly. The study of Chiang, Kuo, and Chen (2009) earlier has indeed already confirmed this distinction. In their experiment, second-year interpreting trainees had significantly shorter average pause duration and lower number of pauses than first-year students in English-to-Chinese sight translation, completely matching our findings. Therefore, we can safely say that the flexibility in language, as well as the resourcefulness, requires time to develop and does not happen "overnight".

Our findings on the influence of experience may have important implications for interpreter training. After receiving training, interpreting students may acquire the know-how to efficiently handle the task of sight translation. But in the meantime, trainees need to know how they can move to the next level by repeatedly trying out different solutions to develop that flexibility in language.

5.4 Intertwined relationships between indices and the skill of chunking, and even reading ahead…

On the face of it, it seems that comparisons aren't possible between experienced interpreters and untrained bilinguals, since the two groups differ in more than one aspect. Indeed, experts easily defeated bilinguals in almost every aspect. However, we still spotted some interesting results.

First of all, the total word counts did not differ between the two groups. This helped dispel the myth among many trainees that conciseness is the ultimate goal. With the same number of words, experts still scored significantly higher than bilinguals regarding both accuracy and fluency. Secondly, while experts required significantly shorter GD, GPT and RRT than bilinguals in silent reading, all reading indices turned out to be non-significant in sight translation. It shows that the additional cognitive load borne by the task still put a large enough weight on experts' shoulders, narrowing the gap between the two groups to within a non-significant range. Even with training and years of experience, this burden could not be reduced to a negligible level. Yet, what's even more fascinating is that we can actually see the trace of reformulation in comparisons regarding FFD. Experts and bilinguals were comparable on this index in both silent reading and sight translation. However, as we have discussed earlier, compared with silent reading, experts lengthened the FFD in sight translation, while bilinguals manifested an opposite trend, corroborating Chen's (2013) speculation that experienced interpreters did more than comprehension during the first pass of reading.

Still, average efficiency during each fixation is only part of the story. Similar length on all reading indices but significantly less total time and number of fixations indicates that experts must have had fewer reading passes. In other words, the real gap might come from the reading pattern during the process of sight translation. Adept at the skill of chunking and flexible in language use, experts could sometimes successfully sight-translate the text or at least make partial attempts at solutions on the first encounter, exempting themselves from having to regress to previous regions for reformulation only in the second reading pass or beyond.

To visualize the reading behavior in sight translation, three cases are provided as examples as follows. The same trial, with the same task sequence and content, was chosen for one member from each group, so we can see clearly how different groups of participants proceeded under the same circumstances.

Figure 7 maps how an expert tackled the fourth trial of Text A with the number of fixations (y-axis) the expert had before uttering each word (x-axis).

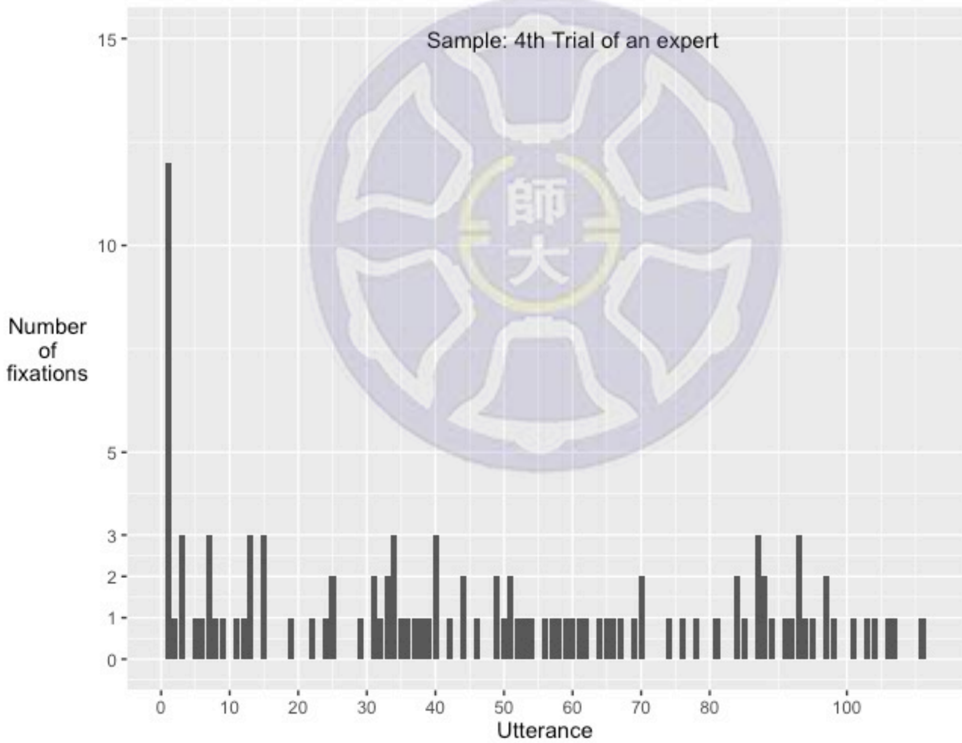


Figure 7 How many fixations experienced interpreters had in each verbal gap: Example of an expert on one trial

As depicted in Figure 7, all reading-ahead spans were limited to three fixations at most, except for the first utterance. After the initial silence for around three seconds

(12 fixations), oral production began. All the way towards the end of this trial, the expert did not dwell on any single part of the text for too long. The number of fixations was always maintained between one and three fixations, equivalent to around one second tops, and at least more than half of the time the expert only fixated once, producing a smooth rendition. This figure is self-evident that it's impossible to only start reformulation in the second reading pass. In other words, some words must have been planned or sight-translated during the first pass.

The reading pattern of a trainee during the same trial is as follows (Figure 8). An initial nine fixations illustrated a fruitful result of training. After roughly a two-second silence, the trainee quickly started. For at least half of the time, the progress was smooth, and only one fixation was needed during each interval. However, the trainee seemed to hit some bumps along the way, so the fixation number went up to between four and seven in order to think of a better translation or to reorganize information, leading to some pauses between one and two seconds. The number of fixations spiked the second time in the middle and was even longer than the silent period in the beginning.

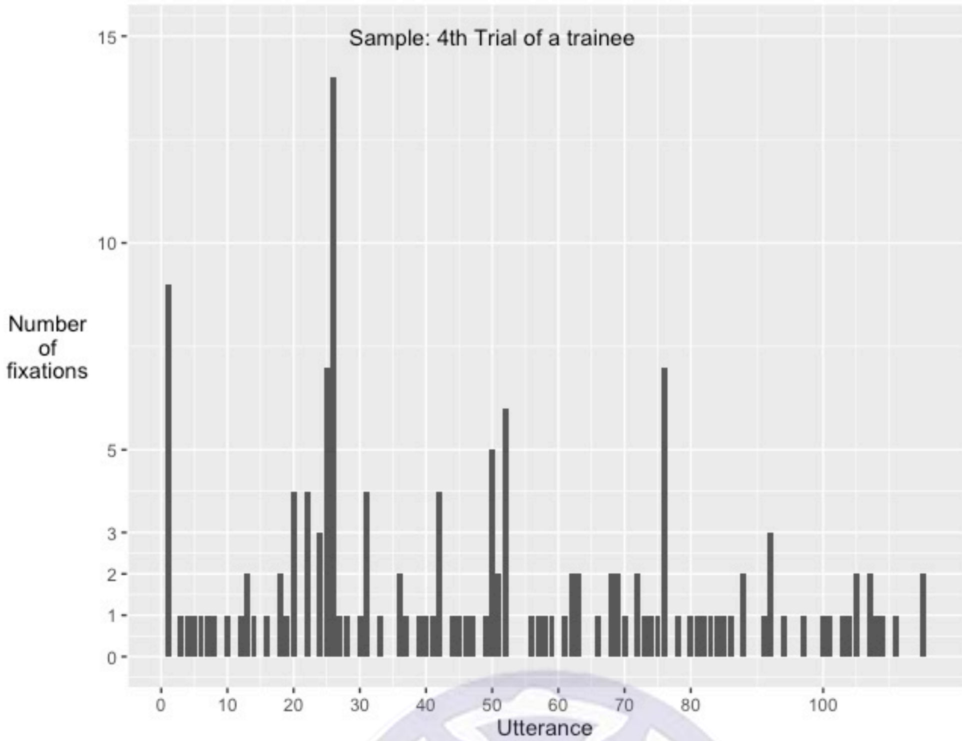


Figure 8 How many fixations interpreting students had in each verbal gap: Example of a trainee on one trial

This clearly indicates that a problem emerged, possibly one that arose due to the fact that the trainee hadn't developed enough elasticity to continue the oral production without an obvious pause. Judging from the length of the initial silence and of pauses that followed, I would suggest that trainees' behavior actually resembled experts'. Therefore, though we argued earlier that a shorter FFD and GD in sight translation for interpreting students signaled a strategy to go back to previous regions to start reformulation as quickly as possible, it is also likely that some reformulation or production were handled within the first reading pass.

Figure 9 shows the reading behavior of a bilingual for the same trial. The bilingual used almost 40 fixations to pre-read the whole trial during the initial silent period, as almost all the bilinguals did. At this point, it is safe to say that for most of the untrained participants in this study, reformulation and production began only after the first reading pass, since first-pass reading indices were also shortened. Once sight translation began, one-fixation intervals were also prevalent, amounting to at least half of all the verbal gaps. The bilingual sometimes ran into difficulties along the way as well. The difference between the trained and the untrained is that the latter had not been taught to chunk messages into shorter segments. Thus, more time was needed when they had to overhaul the content and put each part into the right position in order to sound more natural in the ear of the audience.

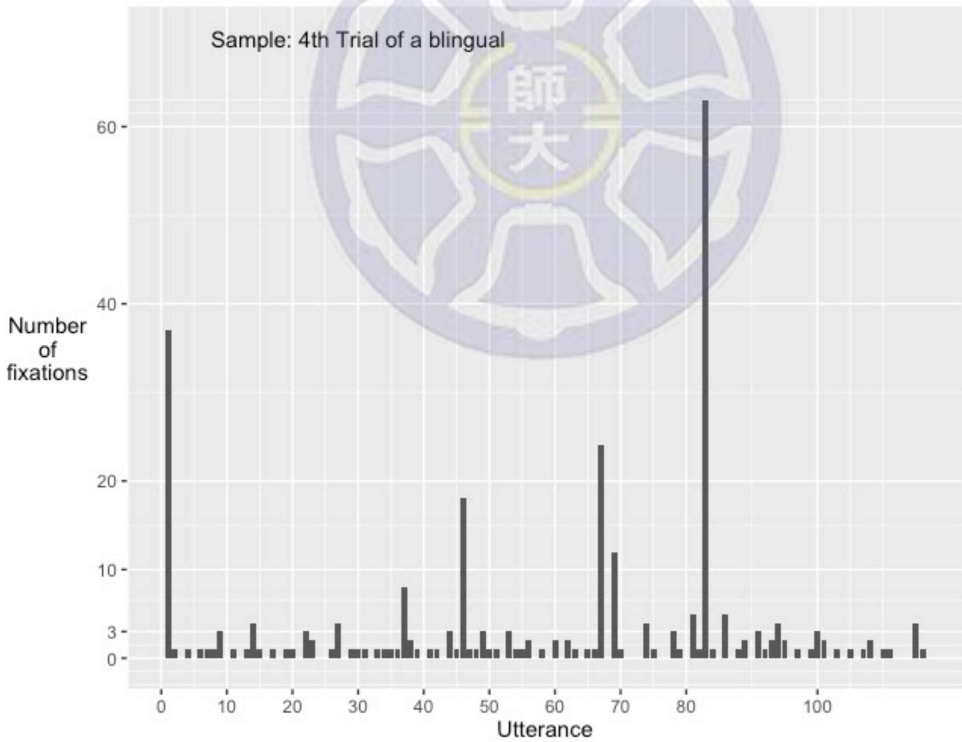


Figure 9 How many fixations untrained bilinguals had in each verbal gap: Example of an average bilingual on one trial

Huang (2011) and Chen (2013) suggested that the first reading pass during sight translation was pure reading for comprehension, as far as interpreting students were concerned. And according to Chen (2013), experienced interpreters seemed to be engaging in multiple tasks rather than simply reading, possibly starting reformulation simultaneously. I would like to stress that interpreting students did in fact show similar tendency to the experts, as a result of training. Perhaps the only difference is the success rate, which might be subject to variation when trainees start to gain more experience.

As for the pattern of reading ahead, without aligning the exact locations of fixations with the verbal output to examine moment-to-moment changes, this study cannot directly affirm Huang's (2011) findings. Nonetheless, we indeed found a noteworthy read-ahead pattern: Experts tried to build their whole oral output upon few quick peeks before uttering each word, and they were adept at the skill of speedy reformulation whenever new information came in. Interpreting students tended to "read ahead" in a similar way with a lesser degree of precision. Therefore, when they were really caught up in problems, longer pauses and more fixations ensued. At the other end of the spectrum, untrained bilinguals did actually "read ahead" as well, just in a different way. Without training, most were used to gleaning the full picture by reading through the text before sight translation began. Further, whenever they ran into problems or finished a complete meaningful unit, bilinguals would then pre-read the next segment, leading to much longer pauses and more fixations.

Su (2013) generally found congruent results with McDonald and Carpenter's (1981) model. The findings in our study match the three-stage model to a certain extent — but with exceptions. As previously discussed, experts and trainees had similar rereading counts in silent reading and sight translation. Coupled with pausing patterns and the number of fixations before each utterance in sight translation, we would again suggest that some reformulation and production might have been completed in the first pass of reading as well. In terms of Hsieh (2014), one of her findings (i.e. experts sometimes finished production during the first reading pass) is only indirectly supported by our results. With a quick start, a fast sight translation speed, much fewer pauses, and the fact that one-fixation per utterance in fact accounting for 75% of the time, we extrapolated that experts could not possibly wait till the second pass to start reformulating and producing output for every chunk; easier or shorter ones must at least have been addressed and solved on the spot during the first encounter. Unfortunately, due to the aim of this study is not to portray the behavior during pauses amid sight translation but to give a brief description of what happens along the way, more detailed comparisons are not possible within the time and resource constraints. Besides, the definition for a pause varies between this study and previous ones. While Su (2013) and Hsieh (2014) took verbal gaps longer than 200 ms as pauses, only those that were above 600 ms were considered as pauses in this study. As previously mentioned, this is due to the fact that pauses below 600 ms were seldom detected by our listeners. Definition aside, around 85-90% of the verbal gaps were over 200 ms in the data under scrutiny in this study. If the same standard was adopted, it would look like that every participant paused after saying each word in such a clumsy way. However, the fluency score the participants received in this study proved otherwise. Therefore, a different definition for a pause was warranted.

5.5 The real deal during pauses

When did participants pause in the midst of the sight translation task? Furthermore, what did they do when pausing? Did training or experience make a difference? Table 33 unveils part of the mystery of the participants' cognitive process. The comparability of the percentage of hesitation pause (and also juncture pause) appears to be implying that experience or training did not help at all at first sight, since the percentage for professionals was even slightly higher than the other two groups. However, once we take into consideration the actual number of pauses, the pattern starts to make more sense: Compared with bilinguals, experienced interpreters only had a little less than half. Bilinguals, though had higher percentage of juncture pause, in reality had much more hesitation pauses as well. Hence, this phenomenon can only serve as an indicator of the difficulties of sight translation, which could still "trap" the professionals in the field of interpreting. In other words, experts were adept at the skill of chunking and reformulation, and most of the time their rendition was so smooth that the audience could not even discern a pause. They were only "caught" once in a while.

Table 33 Features related to pauses[8] (reference table: 21 participants)

Pausing in sight translation (across groups)			
Percentage of HP	NEW=BIL=PRO		
Fixations on PBD units in pause	NEW=BIL=PRO		
Fixations on PBD units in HP	NEW=BIL=PRO		
Fixations on PBD units in JP	NEW=BIL=PRO		
Pause types & Attraction of PBD (within-group)			
	PRO	**NEW**	**BIL**
HP vs. JP	JP=HP	JP<HP	JP<HP
Fixations on PBD units in HP vs. in JP	JP=HP	JP<HP	JP=HP

*PRO = Experienced interpreters; NEW = Interpreting students; BIL = Untrained bilinguals
*HP = Hesitation pause; JP = Juncture pause
*PBD = Principal branching direction

A general high tendency to look at PBD units across groups, especially in hesitation pauses, may attest to the fact that the reversed positions for adjuncts in Chinese and English do in fact lead to greater processing difficulties, as also proven by Viezzi (1989) through a much lower retention rate of the content in languages with relatively contrastive linguistic structures.

A special attention to within-group peculiarities regarding the relative percentage of each pause type and behavior therein may give us more detail on how training or experience actually changed one's behavior. Juncture pauses were significantly fewer than hesitation pauses for trainees and bilinguals, with experts as a group being the only exception. It shows, with all the overwhelming strengths manifested above, experts in general were still able to control themselves in the process, avoiding sounding hesitant too often. As for interpreting students' significantly fewer fixations

8 Numerically speaking, the task placed on the left of any logical symbol is always lower than the one on the right. In addition, statistical significance is represented with “>” or “<”, and “=” stands for non-significance.

on PBD units in juncture pauses, unyielding observance of the chunking principle may by far be the most plausible reason. As we have already mentioned in the last chapter, contrastively positioned PBD units almost never come right after where a juncture pause may possibly appear. Therefore, a set perimeter for reading ahead deprived trainees of the opportunity to choose to look for a specific type of information first. Notwithstanding the speculation, the fixation rate was still around 50%, so we would again assert that PBD units were quite attractive for all groups.

Despite the uniqueness of PBD structures and the attention it drew in this study for all groups of participants, experienced interpreters and interpreting students still managed to limit their processing time on PBD units to a similar level to that on non-PBD units. On the contrary, bilinguals inevitably spent significantly longer time on PBD units in stages of meaning integration and ultimate comprehension. The training of chunking sentences into smaller segments may yet again account for the dichotomous results. Trainees in the English-to-Chinese sight translation class have been repeatedly asked to break their habits of reading through the text. They are instead encouraged to stop reading further once the boundary of a smallest meaningful unit is found and to immediately start rendering, focusing on how to make their output natural and grammatically correct and coherent with the information available at hand. Consequently, no statistical difference was found in any stage between the processing load of non-PBD and PBD units, a result maybe due partly to the fact that PBD units were broken down into smaller and easier segments, or that the PBD structures were tackled as they were predicates, not modifiers, of subjects.

Hsieh (2014) found that some experienced interpreters tended to regress to previous segments where production had already been completed. The author's argument might be a possibility to explain experienced interpreters' behavior in the present study, but unfortunately we are unable to ascertain whether it applies here with the data currently available.

Chapter 6 Conclusion

The current study aims to find out whether reading purposes will change reading behavior, to understand the impacts of training and experience on sight translation respectively, and also to describe the process of sight translation. Experienced interpreters, interpreting students, and untrained bilinguals with comparable linguistic capabilities were invited to participate in the experiment, which entailed tasks of silent reading, reading aloud, and sight translation.

6.1 Reading purposes matter, training has immediate observable influence, but experience still has a role to play

It was found that, in general, reading purposes did change one's reading behavior, but, in the meantime, reading in different tasks still shared some common features. Reading for comprehension was much similar to reading for sight translation in the first pass, both deviating from reading aloud due to its extra burden of vocalization. Moving on to later stages, the cognitive load of reformulation and production for sight translation started to weigh in, leading to a somewhat different pattern from silent reading. However, the above findings need to be interpreted with caution, since similar duration in reading indices does not necessarily represent an identical mechanism or process. A small amount of reduction in FFD for interpreting students and bilinguals and an opposite approach adopted by experienced interpreters in sight translation is just a perfect example. Still, global indices such as total number of fixations and total time clearly outlined the dissimilarities between tasks.

With the data collected in this study, the impacts of training have been successfully identified. Trainees, while identical to bilinguals in every way during silent reading and reading aloud, significantly outperformed the latter group in almost all indices in terms of sight translation, including speed, fluency, accuracy, and mean fixation duration on different local reading indices. Although the gaps in reading indices were not wide enough to reach significance, the lead by trainees was extremely consistent. Comparisons between interpreting students and experienced interpreters, with superior abilities, showed even more amazing influence of training. With formal training for one semester, trainees were already able to catch up with experts in most aspects. In addition to an overall fluency, all local reading indices and global indices, including the pattern of rereading counts, also turned out to be at the same level for the two groups. A look at the processing strategy amid sight translation is also evident that trainees complied with a similar principle: They learned to chunk longer sentences into short segments and attempted at finishing each meaningful unit without regressing back to previous regions again.

However, obstacles that could not be overcome with simply a limited length of training still managed to show the influence of experience. First and foremost, experienced interpreters successfully secured a significantly higher accuracy score. Looking deeper into the process, we found that experts on average could maintain a faster delivery with equal fluency, while observable pauses were kept to a minimum. Even when pauses were inevitable, shorter length and fewer fixations were needed to maneuver around difficulties. What’s more astonishing is that experts on average had significantly fewer fixations (a mere one and a half) every time before they uttered the next Chinese character.

Experienced interpreters and interpreting students both seemed to have adopted the strategy of chunking, tackling the text segment by segment instead of a read-through in advance. Aside from the sample figure provided, we were able to observe the strategy used through a highly converging RRT and TVT between experts and trainees in sight translation and also similar amount of rereading counts between silent reading and sight translation for both groups. Despite the same guiding rule, each group in reality was unique in the way sight translation was carried out. In this regard, trainees were much similar to bilinguals, both speeding up first-pass reading in preparation for a second-pass reformulation or production, while experts had slightly longer FFD and tried to tackle at least part of the reformulation work in the first pass.

Lastly, pause types and where the fixations were targeted during each pause also revealed some traces of training and experience, though the dataset was smaller than the one used to glean all the aforementioned results. Between-group comparisons failed to show significantly lower percentage of hesitation pause for either experts or trainees against bilinguals, and the percentage of fixations on PBD units was also equal across groups. However, experts, as a matter of fact, had the fewest number of hesitation pauses, followed by trainees and then bilinguals, and they were also the only group of participants that managed to show comparable hesitation and juncture pauses. On the other hand, trainees were the only group that fixated significantly less frequently on PBD units in juncture pauses than in hesitation pauses, exhibiting a strict adherence to the chunking skill learned during training.

Last but not least, regardless of pause types, a general high tendency to look at PBD units during pauses across all groups really proved that distinctive linguistic structures in each language indeed could stir up trouble, thereby attracting all participants' attention most of the time. Notwithstanding, in terms of the cognitive load of PBD and non-PBD units in general, experienced interpreters and interpreting students showed no statistical difference in these two types of structures. Untrained bilinguals, on the opposite, had to devote more cognitive resources to processing PBD units to ensure correct interpretation of the message.

6.2 Implications for interpreter training and beyond

Nature or nurture has always been an intriguing topic in the field of interpreting studies. In the current study, all the significant differences between interpreting students and untrained bilinguals in sight translation regarding reading and pausing behavior and also the output assessment have successfully proved that interpreters can actually be trained, and the change in one's behavior is readily visible after a shorter period of training, when compared with the development of expertise. The similarities between the behavior of trainees and of experienced interpreters in sight translation have also attested to the effects of training.

Notwithstanding, some differences remain between trainees and experts, such as the overall accuracy of the output, the number and average length of pauses, and the sight translation speed, which can be explained by a disparity in interpreting experience. This links the implications for interpreter training of this study to Huang's (2011) suggestions. That is to say, the skill of chunking the source text into shorter segments should explicitly be taught in class. Furthermore, trainees should contemplate on how the same text can be paraphrased and sight-translated in various possible ways, so that the flexibility in language use will become a useful weapon, freeing oneself from the confinement of fixed linguistic structures. In this way, the trainees may have a clear goal as to how quality and fluent sight translation can be done, as shown by experienced interpreters in Chen's (2013) and Hsieh's (2014) studies.

Training aside, this study has revealed the basic truth about how interpreters tackle the quasi-simultaneous inflow of written information and outflow of verbal output. From some scholars, the findings here could pave the way for theoretical connections between translation and interpreting, just as the task of sight translation is a hybrid form of the two. Or, what we found here could simply cast some light on more practical issues, such as how sight translation fits into the forever-changing environment of the translation and interpreting industries.

A more far-reaching possibility might be the implication for education of English as a foreign language. Knowing that sight translation skills (mostly, chunking) may improve the efficiency and accuracy when one tackles language, we may introduce sight translation training in language classrooms to help students break complex language structures into easier ones and build up their language flexibility.

6.3 Limitations and suggestions

The limitations of this study are numerous. To begin with, a deviation from the real setting and situation of work was a setback that has also plagued most interpreting studies. Participants had to make extra efforts to hold their heads still, and it was actually reported by some to be a distraction. Secondly, the materials were not strictly controlled as were experiments normally used in the field of psychology. There has always been a tussle between content validity and variable control, especially in the context of interpreting, where real and natural language use needed to be maintained at a certain level. For this reason, this study chose to stick with the original speeches as much as possible, modifying only a small part to adapt the lengthy speeches into three 175-word excerpts. However, the text may still need meticulous control if we wish to truly reveal certain effects in the future. The number of participants, or the amount of data, was yet another factor that limited the explanatory power of this study, especially with regard to the specific comparisons between different types of pauses and fixations on PBD units therein. This is perhaps a hurdle one may inevitably face when experienced interpreters are involved. If time permits, more participants can be invited to ensure enough data are available.

Due to limited time and resource, there are still many questions that haven't been answered. For example, the background information of participants was used in this study to mainly verify if each participant really belonged to the group assigned and whether their qualifications met the requirements for participant selection. Some of the personal profiles haven't been utilized as independent variables, such as the time spent in English-speaking countries, translation or interpreting experience, academic background, age, language exposure, or even language combination. Some of the indices might surprise us with the impacts they have when put in the picture together.

In addition, the process of sight translation was only roughly described in this study. A lack of word-based analysis made moment-to-moment movements of the eye impossible to pin down. Therefore, some findings and extrapolations in previous studies remain unable to corroborate or reject conclusively. Still another possible research direction in the future is that words of different types may pose various challenges for sight translation, and a more detailed look into what happened when the eye move past words of different kinds may help us understand more about the nature of sight translation and even draw inferences based on the knowledge we already have about bilingual reading or underlying cognitive mechanisms of facilitation and inhibition.

Lastly, when more advanced research tools become available in the future, we can further triangulate our findings with different types of data, through different perspectives, and even look for similarities in various modes of interpreting.

References

AIIC (2012). Who can join AIIC? Retrieved from http://aiic.net/p/4046

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, *6*(1), 43-67.

Akinnaso, F. N. (1982). On the differences between spoken and written language. *Language and speech*, *25*(2), 97-125.

Alves, F., Pagano, A., & da Silva, I. (2011). Towards an investigation of reading modalities in/for translation: An exploratory study using eye-tracking data. In O'Brien, S. (Ed.), *Cognitive Explorations of Translation* (pp. 175-196). Bloomsbury Academic.

Angelelli, C. (1999). The role of reading in sight translation. *The ATA Chronicle*, *28*(5), 27-30.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of memory and language*, *26*(6), 658-672.

Berg, T. (2009). *Structure in language: A dynamic perspective*. Routledge.

Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biela-Wolonñciej, A. (2007). A-Vista: New challenges for tailor-made translation types on the example of the recorded sight translation. *Kalbotyra*, *57*, 30-39.

Biq, Y. O. (1995). Chinese causal sequencing and yinwei in conversation and press reportage. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 21, No. 2, pp. 47-60).

Brady, S. A. (1991). The role of working memory in reading disability. *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*, 129-151.

Chafe, W. & Danielewicz, J. (1987). Properties of spoken and written language. Retrieved from http://www.nwp.org/cs/public/download/nwp_file/142/TR05.pdf?x-r=pcfile_d

Chang, H. F. (2008). A study on GITI students' ST difficulty & performance. *SPECTRUM: Studies in Language, Literature, Translation, and Interpretation, 2*, 61-83.

Chen, D. I. (2013). *Differences in comprehension process between experienced and novice interpreters – an eye movement study* (Unpublished M.A. thesis). National Taiwan Normal University, Taipei.

Chen, J. G. (2006a). *Contrastive research & crosslinguistic influence: Some implications for teaching Chinese and English as a second language*. Taipei: Crane.

Chen, J. G. (2006b). Interplay between forward and backward transfer in L2 and L1 writing: The case of Chinese ESL learners in the US. *Concentric: Studies in Linguistics*, *32*(1), 147-196.

Chen, J. G. (2007). The L2 Acquisition of Information Sequencing in Chinese: The case of English CSL learners in Taiwan. *Electronic Journal of Foreign Language Teaching*, *4*(2), 170–191.

Chen, J. G. (2011). *Contrastive analysis and its applications in language pedagogy*. Taipei: Crane.

Chen, W. (2015). Sight translation. In Mikkelson, H. & Jourdenais, R. (Eds.), *The Routledge handbook of interpreting* (pp. 144-153). London: Routledge.

Chernov, G. V. (1994). Message redundancy and message anticipation in simultaneous interpretation. In Lambert, S. & Moser-Mercer, B. (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 139-153). Amsterdam/Philadelphia: John Benjamins.

Chiang, H. K., Kuo, F. L. & Chen, W. C. (2009). Exploring pausing patterns in two-way sight translation. *SPECTRUM: Studies in Language, Literature, Translation, and Interpretation, 5*, 155-167.

Chmiel, A. & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In Way, C., Vandepitte, S., Meylaerts, R., Bartłomiejczyk, M. (Eds.), *Tracks and Treks in Translation Studies* (pp. 189-205). Amsterdam/Philadelphia: John Benjamins.

Christoffels, I. K. (2004). *Cognitive studies in simultaneous interpreting* (Unpublished doctoral dissertation). University of Amsterdam: The Netherlands.

Christoffels, I. K. & de Groot, A. (2005). Simultaneous interpreting: A cognitive perspective. In Kroll, J. F. & de Groot, A. (Eds.), *Handbook of Bilingualism* (pp. 454-479). Oxford: Oxford University Press.

Christoffels, I. K., De Groot, A., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*, 324-345.

Christoffels, I. K., de Groot, A., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition, 6*(3), 201-211.

Chu, C. H. (1998). *A discourse grammar of Mandarin Chinese*. New York: Peter Lang.

Clark, H. H. & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Coltheart, M. (2000). Dual routes from print to speech and dual routes from print to meaning: Some theoretical issues. In Kennedy, A., Radach, R., Heller, D., & Pynte, J. (Eds.), *Reading As a Perceptual Process* (pp. 475-490). Elsevier.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Darò, V., & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: Effects of phonological interference. *Applied Linguistics*, *15*(4), 365-381.

De Groot, A. (2013). Reading. In Grosjean, F. & Li, P. (Eds.), *The Psycholinguistics of Bilingualism* (pp. 73-99). Oxford: Wiley-Blackwell.

De Groot, A., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition, 9*(2), 189-201.

De Groot, A., & Christoffels, I. K. (2007). Processes and mechanisms of bilingual control: Insights from monolingual task performance extended to simultaneous interpretation. *Journal of Translation Studies*, *10*, 17-41.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge Studies in Linguistics. (No. 105). New York: Cambridge University Press.

Diez, D. M., Barr, C. D. & Çetinkaya-Rundel, M. (2015). *OpenIntro statistics (3rd ed.)*. OpenIntro, Inc.

Dijkstra, T., Grainger, J., & van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496-518.

Dijkstra, T., van Jaarsveld, H., & ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51-66.

Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In Lambert, S. & Moser-Mercer, B. (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 155-189). Amsterdam/Philadelphia: John Benjamins.

Dragsted, B. & Hansen, I. G. (2009). Exploring translation and interpreting hybrids: The case of sight translation. *Meta: Translators' Journal, 54*(3), 588-604.

Dragsted, B., Mees, I. M. & Hansen, I. G. (2011). Speaking your translation: Students' first encounter with speech recognition technology. *Translation and Interpreting, 3*(1), 10-43.

Duffy, S.A., Morris, R.K., and Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429-446.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners. *Studies in Second Language Acquisition*, *25*(4), 529-557.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *10*(1), 101-116.

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

Fédération des parents francophones de Colombie-Britannique (n.d.). Bilingualism — Types of bilingualism. Retrieved from http://developpement-langagier.fpfcb.bc.ca/en/bilingualism-types-bilingualism

Fernández, E. (2002). Relative clause attachment in bilinguals and monolinguals. In Heredia, R. R. & Altarriba, J. (Eds.), *Bilingual sentence processing* (pp. 187-215). Amsterdam: North-Holland.

Ford, C. E. (1993). *Grammar in interaction: Adverbial clauses in American English conversations*. New York: Cambridge University Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210.

Frenck-Mestre, C. (1997). Examining second language reading: An on-line look. In Proceedings of the GALA 1997 conference on language acquisition (pp. 474-478).

Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In Heredia, R. R. & Altarriba, J. (Eds.), *Bilingual sentence processing* (pp. 187-216). Amsterdam: North-Holland.

Frenck-Mestre, C. (2005). Ambiguities and anomalies: What can eye movements and event-related potentials reveal about second language sentence processing? In Kroll, J. F. & de Groot, A. (Eds.), *Handbook of bilingualism* (pp. 268-281). Oxford: Oxford University Press.

Erickson, A., Bonet, J., Festinger, N., Framer, I., & Macfarlane, A. G. (2006). Modes of interpreting: Simultaneous, consecutive and sight Translation. *NAJIT position paper*. Retrieved from http://courts.alaska.gov/language/modes-interpreting.pdf.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*(1), 71-99.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*, 263–329.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In Danks, J. H., Shreve, G. M., Fountain, S. B., & McBeath, M. K. (Eds.), *Cognitive processes in translation and interpreting* (pp. 196-214). Thousand Oaks/London/New Delhi: Sage.

Gile, D. (1999). Testing the effort models' tightrope hypothesis in simultaneous interpreting — A contribution. *Hermes*, *23*, 153-172.

Gile, D. (2001). The history of research into conference interpreting: A scientometric approach. *Target, 12*(2), 297-321.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins.

Goldman-Eisler, F. (1972). Segmentation of input in simultaneous translation. *Journal of Psycholinguistic Research*, *1*(2), 127-140.

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

Göpferich, S., Jakobsen, A. L., & Mees, I. M. (Eds.). (2008). *Looking at eyes: Eye-tracking studies of reading and translation processing*. Frederiksberg: Samfundslitteratur.

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences*, *12*(10), 381-387.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67-81.

Grosjean, F. (2001). The bilingual's language modes. In Nicol, J. (Ed.), *One mind, two languages: Bilingual language processing*, (pp. 1-22). Malden, MA: Blackwell.

Her, E. (1997). Binary error analysis of sight interpretation from English into Chinese and its pedagogical implications. *Studies of Interpretation and Translation, 2*, 111-135.

Hernández, A., Fernández, E., & Aznar-Besé, N. (2009). Bilingual sentence processing. In Gaskell, M. G., Gerry, A. (Eds.), *The Oxford handbook of psycholinguistics* (pp. 371-384). Oxford: Oxford University Press.

Ho, C. E. (2013). The review of interpretation research in Taiwan: 2004-2013. *Studies of Translation and Interpretation*, *16*, 111-144.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Psychology Press.

Hsieh, J. H. (2014). *Comparing cognitive processes during pauses between experienced interpreters and novice interpreters: Eye movements in sight translation* (Unpublished M.A. thesis). National Taiwan Normal University, Taipei.

Huang, C. C. (2011). *Tracking eye movements in sight translation — The comprehension process in interpreting* (Unpublished M.A. thesis). National Taiwan Normal University, Taipei.

Huang, L. S. (2009). The potential influence of L1 (Chinese) on L2 (English) communication. *ELT Journal*, *64*(2), 155-164.

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In Hyönä, J, Radach, R., & Deubel, H. (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313-334). Amsterdam: Elsevier.

Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, *48*(3), 598-612.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In Underwood, G. (Ed.). *Eye guidance in reading and scene perception* (pp. 29-53). Oxford: Elsevier.

International Medical Interpreters Association (2009). *IMIA guide on sight translation*. Retrieved from http://www.imiaweb.org/uploads/pages/380_2.pdf.

International Organization for Standardization (ISO). (n.d.). *ISO 13611:2014 (Interpreting — Guidelines for community interpreting)*. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=54082.

Isham, W. P. (2000). Phonological interference in interpreters of spoken-languages: An issue of storage or process? In Dimitrova, B. E. & Hyltenstam, K. (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives* (Vol. 40), (pp. 133-149). Amsterdam/Philadelphia: John Benjamins Publishing.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgment tasks. *Journal of Memory and Language*, *50*(2), 165-181.

Jiang, W. Y. (2009). *Acquisition of word order in Chinese as a foreign language*. Walter de Gruyter.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *Neuroimage*, *20*(2), 693-712.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.

Kambe, G., Rayner, K., & Duffy, S. A. (2001). Global context effects on processing lexically ambiguous words: Evidence from eye fixations. *Memory and Cognition*, *29*(2), 363-372.

Kirkpatrick, A. (1993). Information sequencing in Modern Standard Chinese in a genre of extended spoken discourse. *Text, 13*(3), 423-453.

Klonowicz, T. (1994). Putting one's heart into simultaneous interpretation. In Lambert, S. & Moser-Mercer, B. (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 213-224). Amsterdam/Philadelphia: John Benjamins Publishing.

Koo, T. K. & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Lee, J. (2012). What skills do student interpreters need to learn in sight translation training? *Meta: Translators' Journal*, *57*(3), 694-714.

Lee, T. (2006). A comparison of simultaneous interpretation and delayed simultaneous interpretation from English into Korean. *Meta: Translators' Journal, 51*(2), 202-214.

Levitin, D. J. (2014). *The organized mind: Thinking straight in the age of information overload*. New York: Dutton.

Levy-Schoen, A. (1981). Flexible and/or rigid control of oculomotor scanning behavior. In Fisher, D. F., Monty, R. A, & Senders, J. W. (Eds.), *Eye movements: Cognition and visual perception* (pp. 299-314). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Liao, P. S. (2007). The review of current interpretation research publications in Taiwan. *Studies of Translation and Interpretation*, *10*, 191-219.

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

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Liu, M. H. (2011). Methodology in interpreting studies: A methodological review of evidence-based research. In Nicodemus, B. & Swabey, L. (Eds.), *Advances in interpreting research: Inquiry in action* (pp. 85-119). John Benjamins Publishing.

Liu, M. H., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, *6*(1), 19-42.

Lonsdale, D. (1997). Modeling cognition in SI: Methodological issues. *Interpreting*, *2*(1-2), 91-117.

Macizo, P., & Bajo, M. T. (2004). When translation makes the difference: Sentence processing in reading and translation. *Psicologica: International Journal of Methodology and Experimental Psychology*, *25*(1), 181-205.

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, *99*(1), 1-34.

McDonald, J. L., & Carpenter, P. A. (1981). Simultaneous translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior, 20*(2), 231-247.

Mellinger, C. D., & Hanson, T. A. (2017). *Quantitative research methods in translation and interpreting studies*. London: Routledge.

Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In Carreiras, M., & Clifton, Jr. C. (Eds.), *The on-line study of sentence comprehension* (pp. 15-32). New York: Psychology Press.

Mizuno, A. (2005). Process model for simultaneous interpreting and working memory. *Meta: Translators' Journal*, *50*(2), 739-752.

Moser-Mercer, B., & Setton, R. (2000). The Geneva (ETI) perspective on interpretation research. *Conference Interpretation and Translation*, *2*, 49-56.

National Council on Interpreting in Health Care. (2009). *Sight translation and written translation: Guidelines for healthcare interpreters.* Retrieved from http://www.ncihc.org/assets/documents/publications/Translation_Guidelines_for_ Interpreters_FINAL042709.pdf

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. New York: Cambridge University Press.

O'Grady, W. (2011). Relative clauses: Processing and acquisition. In Kidd, E. (Ed.), *The acquisition of relative clauses: Processing, typology and function* (pp. 13-38). John Benjamins.

Padilla, F., Bajo, M. T., & Macizo, P. (2005). Articulatory suppression in language interpretation: Working memory capacity, dual tasking and word knowledge. *Bilingualism: Language and Cognition, 8*(03), 207-219.

Pöchhacker, F. (2004). *Introducing interpreting studies*. Psychology Press.

Pöchhacker, F. (2008). The turns of interpreting studies. In Hansen, G., Chesterman, A., Gerzymisch-Arbogast, H. (Eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile* (pp. 25-46). Amsterdam/Philadelphia: John Benjamins.

Pöchhacker, F. (2015). Evolution of interpreting research. In Mikkelson, H. & Jourdenais, R. (Eds.), *The Routledge handbook of interpreting* (pp. 62-76). London: Routledge.

Prideaux, G. D. (1989). Text data as evidence for language processing principles: The grammar of ordered events. *Language Sciences*, *11*(1), 27-42.

Prideaux, G. D., & Hogan, J. T. (1993). Markedness as a discourse management device: The role of alternative adverbial clause orders. *Word*, *44*(3), 397-411.

Pym, A. (1992). Translation error analysis and the interface with language teaching. In Dollerup, C. & Loddegaard, A. (Eds.), *The teaching of translation* (pp.279-288). Amsterdam: John Benjamins.

Pym, A. (2009). On omission in simultaneous interpreting: Risk analysis of a hidden effort. In Hansen, G., Chesterman, A., & Gerzymisch-Arbogast, H. (Eds.), *Efforts and models in interpreting and translation research* (pp. 83-105). Amsterdam/Philadelphia: John Benjamins.

Radach, R., & Kennedy, A. (2004). Theoretical perspective on eye movement in reading: Past controversies, current Issues, and an agenda for future research. *European Journal of Cognitive Psychology*, 16(1/2), 3-26.

Ramsay, V. (1987). The functional distribution of preposed and postposed “if” and “when” clauses in written narrative. In Tomlin, R S. (Ed.), *Coherence and grounding in discourse* (pp. 383-408). Amsterdam: John Benjamins.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.

Rayner, K., Cook, A. E., Juhasz, B. J., & Frazier, L. (2006). Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements. *British Journal of Psychology*, *97*, 467-482.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*(2), 31-74.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002). How should reading be taught? *Scientific American*, *286*(3), 70-77.

Rayner, K., Juhasz, B. J. & Pollatsek, A. (2005). Eye movements during reading. In Snowling, M. J. and Hulme, C. (Eds.), *The science of reading: A handbook* (pp. 79-97). Malden, MA: Blackwell.

Rayner, K., & Liversedge, S. P. (2004). Visual and linguistic processing during eye fixations in reading. In Henderson J. M. & Ferreira, F. (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 59-104). New York: Psychology Press.

Rayner, K. & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In Liversedge, S. P., Gilchrist, I. D., & Everling, S. (Eds.), *The Oxford handbook of eye movements* (pp. 751-766). Oxford: Oxford University Press.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr., C. (2012). *Psychology of reading*. Psychology Press.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effects of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290-1301.

Sampaio, G. R. L. (2007). Mastering sight translation skills. *Tradução and Comunicação*, (16), 63-69.

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review*, *19*(2), 185-221.

Schachter, J. (1974). An error in error analysis. *Language Learning*, *24*(2), 205-214.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197-212.

Setton, R. (2001). Deconstructing SI: a contribution to the debate on component processes. *The Interpreters' Newsletter*, *11*, 1-26.

Shreve, G. M., Lacruz, I., & Angelone, E. (2010). Cognitive effort, syntactic disruption, and visual interference in a sight translation task. In Shreve, G. M. & Angelone, E. (Eds.), *Translation and cognition* (pp. 63-84). Amsterdam/Philadelphia: John Benjamins.

Shreve, G. M., Schäffner, C., Danks, J. H., & Griffin, J. (1993). Is there a special kind of "reading" for translation: An empirical investigation of reading in the translation process. *Target*, *5*(1), 21-41.

Stubbs, M. (1983). *Discourse analysis: The sociolinguistic analysis of natural language*. Chicago: University of Chicago Press.

Su, Y. W. (2013). *Cognitive process during pauses in interpreting output: From eye movements in sight translation.* (Unpublished M.A. thesis). National Taiwan Normal University, Taipei.

Tsai, J. L., Lee, C. Y., Tzeng, O. J. L., Hung, D. L., & Yen, N. S. (2004). Use of phonological codes for Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain and Language*, *91*, 235–244.

Tsai, J. L., Yen, M. H., & Wang, C. A. (2005). Eye movement recording and the application in research of reading Chinese. *Research in Applied Psychology*, *28*, 91-104.

Van Hell, J. G., & de Groot, A. M. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431-451.

Van Leerdam, M., Bosman, A. M., & de Groot, A. (2009). When MOOD rhymes with ROAD dynamics of phonological coding in bilingual visual word perception. *The Mental Lexicon*, *4*(3), 303-335.

Van Riemsdijk, H. C., & Williams, E. (1986). *Introduction to the theory of grammar*. The MIT Press.

Viaggio, S. (1995). The praise of sight translation (and squeezing the last drop thereout of). *The Interpreters' Newsletter, 6*, 33–42.

Vonk, W., & Cozijn, R. (2003). On the treatment of saccades and regressions in eye movement measures of reading time. In Hyönä, J., Radach, R., & Deubel, H., (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 291-312). Amsterdam: Elsevier Science.

Wang, Y. F. (2006). The information structure of adverbial clauses in Chinese discourse. *Taiwan Journal of Linguistics*, *4*(1), 49-88.

Weber, W. K. (1990). The importance of sight translation in an interpreter training program. In Bowen, D. & Bowen, M. (Eds.), *Interpreting — Yesterday, today, and tomorrow* (pp. 44-52). Binghamton: State University of New York.

Wu, M. L. (2012). *SPSS operation and application — The practice of quantitative analysis of questionnaire data*. Taipei: Wu-Nan.

Yang, C. L., Perfetti, C. A., & Liu, Y. (2010). Sentence integration processes: An ERP study of Chinese sentence comprehension with relative clauses. *Brain and Language*, *112*(2), 85-100.

Young, W. L. (1982). Inscrutability revisited. In Gumperz J. (Ed.), *Language and social identity* (pp. 72-84). New York: Cambridge University Press.

Appendix I: Experiment materials (with reading difficulty rating sheet)

Text A

1. Environmental issues are woven into the history of the WTO. The founders of this organisation clearly saw that the well-being of habitats, societies, and economies are not separate.

2. Their vision was of global cooperation in trade as a means to unleash growth, alleviate poverty, raise living standards and ensure full employment, while also protecting the environment. There has been a huge increase in international trade since the creation of the WTO.

3. However, as trade has grown, so have concerns about the effects that it may have on the environment. Some aspects of trade actually benefit the environment, through supporting increased efficiency in production, for example. But there is no doubt that there can be harmful effects too, which we must seek to manage and mitigate.

4. Without trade, it is difficult to see how the world can provide for the additional two billion people expected by 2050, and respond to the urgent need to improve living standards, particularly for the poorest. This calls for practical and realistic ways to make trade and environmental policies work together.

Please rate the reading difficulty (not sight translation) of each of the above paragraphs from 1-7, and also give a score (1-7) regarding the general impression of the difficulty level of this text.

1 = really easy

7 = really difficult

Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	General impression

Appendix I (cont.): Experiment materials (with reading difficulty rating sheet)

Text B

1. By welcoming new members, the WTO has expanded the principles and the rule of law which govern economies of all sizes and which stipulate countries of all kinds enjoy the same rights while subject to the same obligations.

2. The WTO plays a crucial role in global governance by setting the global trade rules, monitoring adherence by states to those rules, and helping to resolve disputes between nations when they arise. It further does a great deal to help developing countries of the least-developed to integrate into the global trading system.

3. And we are proud to say that African members constitute a quarter of the entire membership. At the end of 2014 all members recommitted to agreeing on a work programme on the remaining issues of the Doha Development Agenda. This means the big, tough issues of agriculture, services and industrial goods are all back on the table.

4. And it means we have the opportunity to advance negotiations that have been stalled for quite some years. And so we have a big year ahead of us.

Please rate the reading difficulty (not sight translation) of each of the above paragraphs from 1-7, and also give a score (1-7) regarding the general impression of the difficulty level of this text.

1 = really easy

7 = really difficult

Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	General impression

Appendix I (cont.): Experiment materials (with reading difficulty rating sheet)

Text C

1. Trade and environment go hand-in-hand. Seeing the close links between the well-being of habitats, societies, and economies led WTO's founders to embrace firmly the principle of sustainable development. For the same reason, we believe global trade cooperation is vital if we want to achieve both economic and environmental goals.

2. But to make lasting progress on trade and sustainable development, we have to make sure policies in one area go well with policies in other areas. We can't say we have a win-win situation unless one policy creates good outcomes in other areas.

3. We have proved already that there is a positive relationship between trade and the long-term survival of many species. Multilateralism and the rule of law have done a great deal in tackling challenges that reach beyond national borders.

4. And the multilateral trading system is capable of accommodating environmental considerations, including measures that aim to protect and conserve natural resources. Nevertheless, sustainable development remains work in progress still. The goal hasn't been achieved as of today, and we must keep building on current results.

Please rate the reading difficulty (not sight translation) of each of the above paragraphs from 1-7, and also give a score (1-7) regarding the general impression of the difficulty level of this text.

1 = really easy

7 = really difficult

Paragraph 1	Paragraph 2	Paragraph 3	Paragraph 4	General impression

Appendix II: Experiment instruction

Frame 1

您好，本實驗旨在了解人們進行默讀、朗讀、視譯三項任務時眼睛會如何移動以捕捉資訊並理解內容。

本實驗包括3篇正式實驗文章，各175字，皆為世貿組織秘書長的演講摘錄，請您1篇默讀、1篇朗讀、1篇視譯（看著螢幕上的文字以口語翻譯出來）；研究者會於文章出現前提醒您目前要進行哪項任務。文章開始前各有一段練習段落，供您熟悉該篇文章要完成的任務，練習完請按反應盒按鈕進入正式文章，每項任務完成皆會請您回答2題理解測驗，實驗將全程錄音。

圖文說明請見後頁。

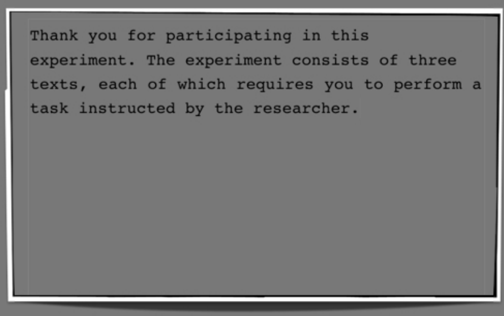
Frame 2

實驗開始，每個畫面的文字出現前，會請您凝視畫面左上角的十字，如下圖：

Appendix II (cont.): Experiment instruction

Frame 3

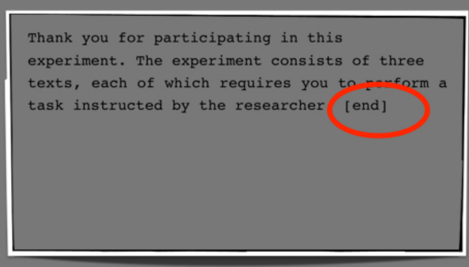
接著練習語料便會出現，如下圖：



此時研究者會告知您需要進行的任務（默讀、朗讀或視譯），練習完按反應盒按鈕便進入正式實驗文章。

Frame 4

每篇正式文章皆分成數個畫面，最後一面的結尾有 [end] 的字樣，請您完成後凝視 [end]，代表整篇文章完成，如下圖：



凝視完 [end] 請按反應盒按鈕進入理解測驗。

Appendix II (cont.): Experiment instruction

Frame 5

Frame 6

Appendix III: Participant consent form

研究同意書

感謝您參與本項研究，探討英進中視譯的認知歷程：以眼動為指標的初探研究。我是何承恩，目前為國立臺灣師範大學翻譯研究所博士候選人。此研究希望能夠透過眼球移動了解視譯當下口譯員一邊閱讀一邊口譯的認知歷程，以及為了視譯進行的閱讀與一般默讀資料時的閱讀有何異同。

如果您同意參加本研究，我會邀請您進行實驗並填寫背景資料問卷。總時間約需 60-80 分鐘，過程將會以眼動儀偵測您的眼球移動，同時錄音。此研究沒有侵入式的風險，錄音內容與背景資料會以代碼註記，不會顯示您的個人資料。一切資料將由研究者妥善保管，非經您同意不會公開。未來論文內容提及研究對象時，會以代號稱呼，絕不會洩露任何個人資料。

如果您現在有任何問題，歡迎直接詢問。之後有任何問題，也可以透過電話 0912711721 或 email: callcallpp@gmail.com 與我聯絡。或者您也可以與我的指導教授陳子瑋教授聯絡，email: d9425@ntnu.edu.tw。您需要的話，可保留一份同意書。

在以下欄位簽名，就表示您已閱讀過以上的說明並同意參與。如果您之後改變心意，只要告訴我就可以隨時退出本研究。

參與者簽名處：　　　　　　　　　　　　　　日期：

研究者簽名處：　　　　　　　　　　　　　　日期：

Appendix IV: Background questionnaire

語言背景問卷與自評量表

本問卷與自評量表旨在了解個人語言能力相關背景以及對於自身語言能力之看法，僅供研究者觀察語料現象時對比使用，不作任何其他用途，請放心填答。

您好，接下來要請您回答一些關於中英文語言學習的經驗，並簡單評量自己的語言能力，最後將詢問您日常使用或接觸翻譯的概況。謝謝您填寫這份問卷。

本問卷包括30個問題。

語言經歷

本大題旨在了解您的語言學習背景以及日常生活中接觸/使用各種語言的概況。

受試者編號 *

請在此填寫您的回應:

[ ]

請填入研究者聯絡您時提供的**實驗編號**

性別 *

以下選項請只選擇一個:

○ 女
○ 男

年齡 *

以下選項請只選擇一個:

○ 18-25
○ 26-33
○ 34-41
○ 42-49
○ 50以上

Appendix IV (cont.): Background questionnaire

請問你的最高學歷是？【需取得畢業證書】

*

以下選項請只選擇一個:

- ○ 大專
- ○ 碩士
- ○ 博士
- ○ 其它 [ ]

請問你就讀的系所為何？【不管您已畢業或仍舊在學，都請您勾選並註明狀態，再簡略填寫系所名稱。例：碩士。在學，外文研究所】 *

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 大專 [ ]
- ☐ 碩士 [ ]
- ☐ 博士 [ ]
- 其他: [ ] [ ]

在學、畢業皆請勾選

請按照習得先後順序列出你所學過的語言：【在左邊空格打v，在右邊空格填1,2,3,4...】

*

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 中文 [ ]
- ☐ 英文 [ ]
- ☐ 日文 [ ]
- ☐ 法文 [ ]
- ☐ 韓文 [ ]
- ☐ 西班牙文 [ ]
- 其他: [ ] [ ]

若雙母語者同時開始學習兩種語言，請都填1。

Appendix IV (cont.): Background questionnaire

請按照學習或接觸語言的時間長短列出你接觸過的語言。【在左邊空格打v，右邊空格依時間長至短填1,2,3,4...】

*

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 中文
- ☐ 英文
- ☐ 日文
- ☐ 法文
- ☐ 韓文
- ☐ 西班牙文
- 其他:

若兩種語言學習時間一樣長，可填入相同數字。

請問你目前生活中使用下列語言的百分比各是多少？【請填多少%，若兩種語言差不多可各填50%，全部總和=100 %】

*

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 中文
- ☐ 英文
- ☐ 日文
- ☐ 法文
- ☐ 韓文
- ☐ 西班牙文
- 其他:

Appendix IV (cont.): Background questionnaire

請問閱讀時，你以哪種語言的閱讀材料為主？【請填多少%，所有項目加起來=100 %】

*

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 中文
- ☐ 英文
- ☐ 日文
- ☐ 法文
- ☐ 韓文
- ☐ 西班牙文
- 其他:

若兩種語言比例相等，可各填50%

請問聊天時，你使用下列各種語言的比例？【請填多少%，所有項目加起來=100 %】

*

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 中文
- ☐ 英文
- ☐ 日文
- ☐ 法文
- ☐ 韓文
- ☐ 西班牙文
- 其他:

若兩種語言使用比例相同，可各填50%

Appendix IV (cont.): Background questionnaire

請問你是否曾居住於全英語環境？如果有，請勾選時間長短，並簡述原因。【如移民、出差、求學等等，但旅遊不算】

*

以下選項請只選擇一個:

- ○ 無
- ○ 未滿1年
- ○ 1年以上-未滿3年
- ○ 3年以上-未滿5年
- ○ 5年以上-未滿7年
- ○ 7年以上-未滿9年
- ○ 9年以上

請說明您的選擇:

本題「全英語環境」意指日常生活以英語為主要溝通語言。使用英語的比例也遠大於其他語言。

Appendix IV (cont.): Background questionnaire

請問你持有哪些語言證照？【請勾選持有的證照，並於右欄輸入分數】 *

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

- ☐ 無
- ☐ TOEFL（CBT）
- ☐ TOEFL（iBT）
- ☐ IELTS
- ☐ TOEIC
- ☐ GEPT
- ☐ BULATS

其他:

請於各欄填入證照的分數，分析時研究者會將一定範圍內的分數歸類為同一群組。僅當作背景變項使用，不帶任何價值判斷，請放心填答。

Appendix IV (cont.): Background questionnaire

語言水平

本大題請您概略評量自己的語言能力，請以自己各項能力為比較基準即可，不需與其他同儕比較。

請問你幾歲開始說中文呢？

*

以下選項請只選擇一個:

- ○ 0-5歲(學齡前-幼稚園）
- ○ 6-12歲(國小)
- ○ 13-15歲(國中)
- ○ 16-18歲(高中)
- ○ 19歲以上(大學之後)
- ○ 其它 [ ]

請問你幾歲開始讀中文呢？

*

以下選項請只選擇一個:

- ○ 0-5歲(學齡前-幼稚園)
- ○ 6-12歲(國小)
- ○ 13-15歲(國中)
- ○ 16-18歲(高中)
- ○ 19歲以上(大學之後)
- ○ 其它 [ ]

Appendix IV (cont.): Background questionnaire

在**1-10**的分數範圍內，請問你給自己的中文聽力理解幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

在**1-10**的分數範圍內，請問你給自己的中文口說能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Appendix IV (cont.): Background questionnaire

在**1-10**的分數範圍內，請問你給自己的中文閱讀能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

在**1-10**的分數範圍內，請問你給自己的中文寫作能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Appendix IV (cont.): Background questionnaire

請問你幾歲開始說英文呢？

*

以下選項請只選擇一個:

- ○ 0-5歲(學齡前-幼稚園）
- ○ 6-12歲(國小)
- ○ 13-15歲(國中)
- ○ 16-18歲(高中)
- ○ 19歲以上(大學之後)
- ○ 其它 [ ]

請問你幾歲開始讀英文呢？

*

以下選項請只選擇一個:

- ○ 0-5歲(學齡前-幼稚園)
- ○ 6-12歲(國小)
- ○ 13-15歲(國中)
- ○ 16-18歲(高中)
- ○ 19歲以上(大學之後)
- ○ 其它 [ ]

在**1-10**的分數範圍內，請問你給自己的<u>英文聽力</u>理解幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Appendix IV (cont.): Background questionnaire

在**1-10**的分數範圍內，請問你給自己的英文口說能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

在**1-10**的分數範圍內，請問你給自己的英文閱讀能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Appendix IV (cont.): Background questionnaire

在**1-10**的分數範圍內，請問你給自己的<u>英文寫作</u>能力幾分呢？ *

以下選項請只選擇一個:

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10

Appendix IV (cont.): Background questionnaire

口譯學習與工作

本大題旨在了解您學習口譯的狀況，以及您的工作使用/接觸口筆譯的概況。所有資訊僅做為本研究之背景變項，以利區分組別，不會用於其他任何用途，請您放心填答。

請問你是否修習過下列口譯課程？ *

請選擇所有符合條件的:

- ☐ 無
- ☐ 學校正規課程
- ☐ 推廣班
- ☐ 其它: ____________

承上題，請問修習口譯課程的時間符合下列哪個選項？【若為學校正規課程，請勾選幾學期；若為推廣班，煩請勾選「其他」並填入大約時數】 *

請選擇所有符合條件的:

- ☐ 一學期
- ☐ 二學期
- ☐ 三學期
- ☐ 四學期
- ☐ 五學期以上
- ☐ 其它: ____________

承上，請問你修習的口譯課程中，包含多久的視譯課程？【若為學校正規課程，請勾選幾學期；若為推廣班，請勾選其他並填入大約時數】 *

請選擇所有符合條件的:

- ☐ 無
- ☐ 一學期
- ☐ 二學期
- ☐ 三學期以上
- ☐ 其它: ____________

Appendix IV (cont.): Background questionnaire

現職：【若僅有一項職業，請勾選後填**1**；若複選，請依主次順序分別填入**1,2,3,4...**】 *

Comment only when you choose an answer.

請選擇所有符合的選項並加以說明:

☐ 口譯
☐ 筆譯
☐ 教師
☐ 學生
☐ 語言工作者
其他:

若有需要可於右欄補充說明

請問你至今為止，口譯工作的<u>總天數</u>約莫幾天？（單場連續超過**3**小時以**1**天計，若有兩個半天以內的場次，也可合併算為**1**天） *

以下選項請只選擇一個:

○ 無
○ 1-29天
○ 30-59天
○ 60-89天
○ 90-119天
○ 120-149天
○ 150天以上

僅供受試者分組使用，無關市場現況調查，請放心填答。

請問你至今為止，筆譯工作的<u>總字數</u>約莫多少？ *

以下選項請只選擇一個:

○ 無
○ 未滿20萬字
○ 20萬字-未滿40萬字
○ 40萬字-未滿60萬字
○ 60萬字-未滿80萬字
○ 80萬字-未滿100萬字
○ 100萬字以上

僅供受試者分組使用，無關市場現況調查，請放心填答。

Appendix V: Paired samples test tables for within-group comparisons

Paired Samples Test

Experienced interpreter	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
FFD_SR_BASE - FFD_SR_PBD	-16.407435	21.661663	9.687390	-43.303943	10.489072	-1.694	4	.166
FFD_RA_BASE - FFD_RA_PBD	13.598903	17.395961	7.779710	-8.001035	35.198841	1.748	4	.155
FFD_ST_BASE - FFD_ST_PBD	-7.317080	33.557150	15.007214	-48.983784	34.349625	-.488	4	.651
GD_SR_BASE - GD_SR_PBD	-11.985742	28.867260	12.909831	-47.829180	23.857696	-.928	4	.406
GD_RA_BASE - GD_RA_PBD	6.894659	31.362645	14.025801	-32.047208	45.836526	.492	4	.649
GD_ST_BASE - GD_ST_PBD	-11.890085	20.544804	9.187916	-37.399828	13.619659	-1.294	4	.265
GPT_SR_BASE - GPT_SR_PBD	-56.968014	69.937062	31.276805	-143.806346	29.870317	-1.821	4	.143
GPT_RA_BASE - GPT_RA_PBD	-7.693308	77.156669	34.505511	-103.495966	88.109350	-.223	4	.834
GPT_ST_BASE - GPT_ST_PBD	-36.833370	96.212434	43.027509	-156.296886	82.630146	-.856	4	.440
RRT_SR_BASE - RRT_SR_PBD	22.183632	55.351822	24.754087	-46.544733	90.911997	.896	4	.421
RRT_RA_BASE - RRT_RA_PBD	-30.854301	73.714560	32.966154	-122.383016	60.674415	-.936	4	.402
RRT_ST_BASE - RRT_ST_PBD	-75.962794	108.539016	48.540123	-210.731782	58.806194	-1.565	4	.193
TVT_SR_BASE - TVT_SR_PBD	-5.243740	37.137361	16.608333	-51.355864	40.868383	-.316	4	.768
TVT_RA_BASE - TVT_RA_PBD	-9.902371	41.500033	18.559379	-61.431469	41.626726	-.534	4	.622
TVT_ST_BASE - TVT_ST_PBD	-82.604455	82.390877	36.846320	-184.906240	19.697330	-2.242	4	.088

Appendix V (cont.): Paired samples test tables for within-group comparisons

Paired Samples Test

Interpreting student	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
FFD_SR_BASE - FFD_SR_PBD	-2.687835	19.280544	6.426848	-17.508173	12.132503	-.418	8	.687
FFD_RA_BASE - FFD_RA_PBD	7.120842	27.607841	9.202614	-14.100423	28.342107	.774	8	.461
FFD_ST_BASE - FFD_ST_PBD	-6.309665	19.422362	6.474121	-21.239014	8.619684	-.975	8	.358
GD_SR_BASE - GD_SR_PBD	-14.999132	29.719428	9.906476	-37.843507	7.845243	-1.514	8	.168
GD_RA_BASE - GD_RA_PBD	-12.857054	40.552848	13.517616	-44.028733	18.314625	-.951	8	.369
GD_ST_BASE - GD_ST_PBD	-24.398817	41.911528	13.970509	-56.614869	7.817235	-1.746	8	.119
GPT_SR_BASE - GPT_SR_PBD	-13.635382	62.640262	20.880087	-61.784950	34.514186	-.653	8	.532
GPT_RA_BASE - GPT_RA_PBD	-13.831864	72.244641	24.081547	-69.364011	41.700283	-.574	8	.582
GPT_ST_BASE - GPT_ST_PBD	5.467726	121.420421	40.473474	-87.864272	98.799724	.135	8	.896
RRT_SR_BASE - RRT_SR_PBD	25.388945	92.241978	30.747326	-45.514516	96.292406	.826	8	.433
RRT_RA_BASE - RRT_RA_PBD	-.019115	133.803355	44.601118	-102.869479	102.831248	.000	8	1.000
RRT_ST_BASE - RRT_ST_PBD	-21.792293	56.256095	18.752032	-65.034555	21.449970	-1.162	8	.279
TVT_SR_BASE - TVT_SR_PBD	-11.207231	85.572413	28.524138	-76.984010	54.569548	-.393	8	.705
TVT_RA_BASE - TVT_RA_PBD	-32.435837	86.525466	28.841822	-98.945197	34.073524	-1.125	8	.293
TVT_ST_BASE - TVT_ST_PBD	-36.643310	73.208745	24.402915	-92.916533	19.629912	-1.502	8	.172

Appendix V (cont.): Paired samples test tables for within-group comparisons

Paired Samples Test

Untrained bilingual	Paired Differences					t	df	Sig. (2-tailed)
				95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
FFD_SR_BASE - FFD_SR_PBD	-14.035312	15.974133	6.037655	-28.808921	.738296	-2.325	6	.059
FFD_RA_BASE - FFD_RA_PBD	17.458877	27.330534	10.329971	-7.817651	42.735405	1.690	6	.142
FFD_ST_BASE - FFD_ST_PBD	-10.529591	21.266729	8.038068	-30.198035	9.138852	-1.310	6	.238
GD_SR_BASE - GD_SR_PBD	-28.867339	27.881477	10.538208	-54.653404	-3.081274	-2.739	6	.034
GD_RA_BASE - GD_RA_PBD	-5.110202	32.065438	12.119596	-34.765786	24.545382	-.422	6	.688
GD_ST_BASE - GD_ST_PBD	-40.873096	22.179120	8.382920	-61.385361	-20.360830	-4.876	6	.003
GPT_SR_BASE - GPT_SR_PBD	-37.612576	83.829109	31.684425	-115.141571	39.916419	-1.187	6	.280
GPT_RA_BASE - GPT_RA_PBD	18.869470	47.085831	17.796771	-24.677661	62.416601	1.060	6	.330
GPT_ST_BASE - GPT_ST_PBD	-36.694272	112.205636	42.409744	-140.467177	67.078633	-.865	6	.420
RRT_SR_BASE - RRT_SR_PBD	-64.991272	63.333245	23.937717	-123.564754	-6.417789	-2.715	6	.035
RRT_RA_BASE - RRT_RA_PBD	-35.731699	59.881003	22.632892	-91.112391	19.648992	-1.579	6	.165
RRT_ST_BASE - RRT_ST_PBD	-88.353317	85.015135	32.132701	-166.979203	-9.727431	-2.750	6	.033
TVT_SR_BASE - TVT_SR_PBD	-43.978446	59.139910	22.352785	-98.673740	10.716849	-1.967	6	.097
TVT_RA_BASE - TVT_RA_PBD	-10.723548	34.183815	12.920267	-42.338304	20.891207	-.830	6	.438
TVT_ST_BASE - TVT_ST_PBD	-99.006129	75.808343	28.652860	-169.117152	-28.895105	-3.455	6	.014