

Chen-En Ho* and Jie-Li Tsai

Reader differences in navigating English–Chinese sight interpreting/translation

<https://doi.org/10.1515/psicl-2024-0091>

Received November 18, 2024; accepted August 28, 2025

Abstract: Reading is key in sight interpreting/translation (SiT), a task in this study involving reading and orally rendering text at one's own pace in a diplomatic interpreting scenario. However, little attention is given to how different reading processes are used. This study bridges this gap by investigating SiT reading processes, using silent reading (SR) and reading aloud (RA) for comparison to understand how reading varies between tasks and participants. Experienced interpreters, interpreting trainees, and untrained bilinguals were recruited to conduct SR, RA, and SiT. Their eye movement data underwent cluster analysis based on fixation duration and saccade length plus direction. Five distinct reading processes have been identified – skimming, rauding, two levels of problem-solving, and anchoring. While the overall reading pattern is similar, nuanced differences tell groups and tasks apart. Due to the limitation of space, this paper only reports findings centring on the participants. Significant differences exist only between the trained (i.e., interpreters and trainees combined) and untrained cohorts in three processes, namely skimming, rauding (normal reading), and problem-solving, almost exclusively in SiT. Our findings attest to the multifaceted SiT reading processes and offer an alternative account to associating fixation duration solely with cognitive load, helping us better understand SiT reading.

Keywords: sight interpreting/translation; eyetracking; reading process; reading pattern; cluster analysis

1 Introduction

This study aims to investigate the use of various reading processes in sight interpreting/translation (SiT), including scanning for, recognising, and decoding words, skimming for ideas, and integrating meaning derived from individual words while

***Corresponding author: Chen-En Ho**, Queen's University Belfast, Belfast, Northern Ireland, E-mail: c.ho@qub.ac.uk <https://orcid.org/0000-0001-7321-2029>

Jie-Li Tsai, National Chengchi University, Taipei, Taiwan, E-mail: jltsai@nccu.edu.tw <https://orcid.org/0000-0001-5413-4889>

considering the context to understand the source text in preparation for delivery in a different language. Silent reading (SR) and reading aloud (RA) are used as reference points to compare how reading processes (and hence strategies) differ. SiT is traditionally more often called sight translation, but the name has been challenged for being non-indicative of the nature of this task (Čeňková 2015). Meanwhile, other names also permeate the research landscape, including sight interpreting, simultaneous translation, *prima vista*, and so on, making it difficult to holistically understand the nature and specificities of the same task in different scenarios and settings (Ho 2022). Hence, we use SiT as an inclusive term to signal the varied forms and contexts in which the same task could happen and then specify the actual mode and language pair involved, in hopes of creating an “hypernym” under which relevant studies can be easily identified and compiled to facilitate discussion and knowledge consolidation (more about the scattered literature and names see Gorszczyńska 2020; Ho 2022; for possible configurations of SiT, see Ho and Xiao 2024; Li 2014). In our study, the SiT task requires the participants to interpret English (L2) speech scripts into Chinese (L1) in a simulated World Trade Organization (WTO) meeting without giving them prior access to the material, i.e., unprepared SiT used in a formal, interpreter-mediated setting for real-time communication.

The role of reading in SiT cannot be emphasised more in the literature as the core, if not the only, method to access the source message for interpreting. Scholars and trainers have long commented on how reading affects SiT performance (Nilsen and Monsrud 2015), and the importance of speed reading and reading ahead (Agri-foglio 2004; Jiménez Ivars 2008; Lee 2012). Previous studies have also accentuated that successful SiT is contingent on “analytical reading and...text analysis, speed reading, identification of main ideas and their links, and concentration” (Li 2015: 173). While earlier SiT studies tend to make suggestions based on product-oriented analysis or practitioners’ reflections (Xiao et al. 2023), the growing popularity of eyetracking technology in SiT studies (especially from 2017 onwards, see Hu et al. 2022) has provided a unique opportunity to observe the cognitive processes of SiT. Numerous studies have explored reading-related issues, be it the stages of SiT and their functions (e.g., Lijewska et al. 2022; McDonald and Carpenter 1981), reading of syntactic asymmetric structures (e.g., Chmiel and Lijewska 2019; Ma et al. 2021; Ma et al. 2022), or the comparisons between SiT reading and other reading tasks (e.g., Alves et al. 2011; Jakobsen and Jensen 2008; Macizo and Bajo 2004; 2006; Ruiz and Macizo 2019; Ruiz et al. 2008). Nonetheless, most SiT studies refer to reading as one single activity (see Section 2.1) and do not differentiate between the nuances of various types of reading. This is an important knowledge gap that the current study aims to bridge, as research in other disciplines has repeatedly shown that reading consists of multiple distinct processes (see Section 2.2), but the same discussion has not yet happened in SiT studies. Deepening our understanding of the nature of SiT

reading and how many processes are enacted in such a complex communicative activity as interpreting have important implications for the development of interpreting studies, as well as interpreter education, as the findings will inform more nuanced description and teaching of how reading can be done in SiT than broad-brush recommendations, such as speed reading and reading ahead, albeit practical and truthful. Ho and Xiao (2024) explore the reading processes of trainees and have preliminarily identified the potential existence of multiple processes, but the findings and discussion are solely based on descriptive statistics. The current study aims to continue the same line of research using more rigorous methods (albeit still being exploratory in nature), including cluster analysis and inferential statistical analysis, to understand the reading processes in SiT. Furthermore, two reading-related tasks and two other groups of participants are included to help us better understand the nature of SiT reading. Accordingly, two research questions are put forward:

1. What are the reading processes involved in SiT, and what does the reading pattern of SiT look like, when compared with SR and RA?
2. How do participants from various backgrounds exhibit different or similar reading processes when performing SR, RA, and SiT?

2 Reader differences in SiT and the multiplicity of reading processes

This section reviews SiT studies that compare reading behaviours between participants with different backgrounds, such as trainees, practitioners, and untrained bilinguals. As we finish Section 2.1 below, we shall see clearly that reading in previous studies has not been further differentiated. To explore this under-researched area, we can draw on the insights from reading psychology (see Section 2.2) to enrich our understanding of SiT reading.

2.1 Reader differences and “uniform” representation of reading in SiT studies

Jakobsen and Jensen (2008) examine six professional translators (including one interpreter) and six translation students' eye behaviour in four tasks: (1) normal reading, (2) reading for translation, (3) reading during L2-L1 SiT, and (4) reading while translating (p. 106). Task-wise, the difference was non-significant for gaze time (i.e., duration of all fixations), and partly significant for mean fixation duration (MFD). Nonetheless, fixation frequency significantly differentiates between the tasks

and groups. On average, practitioners complete each task faster than the students. However, the difference comes from fewer instead of shorter fixations, and the two groups are parallel regarding gaze time. Alves et al. (2011) also look at both fixation count and duration of six translators and six translation students, who conducted L2-L1 reading for (1) comprehension, (2) summarising, and (3) SiT. While MFD turns out to be statistically significantly longest for SiT, the trend is the opposite for fixation count. Interestingly, both indicators show that the students resemble the professionals in their reading behaviour.

Based on the above two studies, Wang and Yan (2018) examine six professional interpreters' and six interpreting students' behaviour during L2-L1 reading for (1) comprehension, (2) summarising, (3) translation, and (4) SiT. As task time and fixation count increase in tandem with task complexity (except between Task 2&3), MFD presents a slightly different picture: SiT is significantly longer, while the others are statistically identical. Descriptive statistics show that professionals tend to use less time and have fewer fixations and shorter MFD (for which SiT is an exception in that the professionals' is actually longer). Later, He and Wang (2021) again examine L2-L1 SiT, using eight professional interpreters and eight interpreting students. Task-wise, similar to Jakobsen and Jensen's (2008) results, there is no between-group difference in total fixation duration (or gaze time), but professionals do use significantly fewer fixations than their counterparts. This study further delves into local areas of interest (AOIs), i.e., low-frequency words/metaphors, which really showcase the between-group differences: the trainees' numbers for total fixation duration, fixation count, regression time (i.e., how long the participant looks back at the AOIs), and the number of regressions all significantly exceed the other group.

Findings from the above studies converge on several fronts and show that professional experience is reflected in the task time and fixation count, but not (mean) fixation duration nor gaze time. Furthermore, details are where the expertise shines: in local problem triggers such as low-frequency words, professionals outpace trainees and re-read less. Complementing the above observations, Chmiel and Lijewska (2019) compare the performance of professional interpreters against interpreting trainees in L2-L1 SiT (but note that individual sentences were used instead of full texts). On average, professionals gaze away less than trainees, especially in objective-relative sentences, indicating stronger tolerance of visual interference (see also Jakobsen and Jensen 2008). Gaze time and task time again demonstrate the expertise advantage. That said, quite some similarity is still noted, including working memory capacity and the strategies used to chunk object-relative sentences, for which "professionals were not more autonomous in their reformulation choices" (Chmiel and Lijewska 2019: 392).

Until now the research design sheds light on the effects of experience (although earlier studies tend to use translators and later ones interpreters). Chmiel and Mazur

(2013) slightly deviates from this paradigm in comparing the SiT performance of ten interpreting students with one year of training against eight with two years of training in L1-L2 SiT. The design thus *indirectly* looks at exposure to the task in training (as there were no dedicated SiT training but rather scattered as an activity in class), not professional contexts. It turns out an additional year does not yield significant impact on task time, fixation count, and the processing of local AOIs including sentences with specific structures and low-frequency terms. The similarity even extends to the fixation count in the warm-up session, in which participants perused the material for 10 s, although less experienced students embraced a different reading approach and scanned to cover larger ground.

Some other studies, using the same comparative paradigm, look beyond the whole task of SiT and drill down into different reading stages. McDonald and Carpenter (1981) asked two professional translators and two untrained bilinguals to perform L2-L1 SiT using garden-path sentences embedded in context and report that the first reading pass is generally normal reading, followed by reformulation (second pass), and error rectification (following passes) if the initial understanding of the ambiguous elements seems erroneous. The reading (speed) of professionals and amateurs were largely (statistically) similar, with one exception: one amateur read unusually fast in the first pass (seemingly adopting a different approach); interpreting speed was also comparable across groups. The between-group difference therefore seems to be quality (accuracy), not speed.

The above findings are mostly supported by later studies that find statistical similarities between normal and SiT reading in early stages, which include decoding words and deriving meaning from them, regardless of SiT training or professional interpreting experience. Ho et al. (2020) tease out the effects of training using 18 interpreting students and 18 untrained bilinguals in three L2-L1 tasks – (1) SR, (2) RA, and (3) SiT. Results for task time, fixation count, and mean fixation duration attest to the comparable language proficiency of the two groups in SR, while in RA the trainees use significantly fewer fixations, although results for the other two indicators remain comparable. Contrarily, the trainees exhibit statistically significantly different behaviour, completing SiT in less time and fewer fixations, rendering MFD the only exception, all the while achieving significantly better accuracy and delivery scores. Interestingly, local word-based analysis presents a counterintuitive picture: the two groups fail to differ in both early and later-stage reading processes in all three tasks.¹ Using the same experimental setup, Ho (2021) compares the performance of 17 experienced interpreters and 18 interpreting students in SiT. The professionals turn out to use statistically similar task time, fixation counts, and MFD as the trainees,

¹ The indicators used include first fixation duration, gaze duration, go-past time, re-reading time, and total viewing time.

scoring significantly higher on accuracy but comparable on delivery. Word-based reading analysis also confirms the between-group similarities in the first and non-first pass reading.

From a quick combing of the SiT literature, three observations emerge: (1) the goal (and reading instructions) differentiates SiT reading (partly) from other reading-related tasks; (2) professionals differ from trainees in aggregate indicators, e.g., total fixation duration/frequency, but rarely in mean fixation duration or even word-based indices such as first fixation duration and gaze duration – in other words, the *de facto* reading behaviour bears much resemblance; (3) there is only one form of reading, represented (almost) solely by duration.

2.2 Insights from reading research

Numerous studies have actually found that reading is multi-faceted, consisting of several reading processes (Carver 1990; Olivier et al. 2022; Simola et al. 2008). While the conclusion remains that purpose affects reading behaviour, acknowledging the multiplicity of reading is important. An obvious benefit is that we can better understand why fixation duration varies more and sometimes increases dramatically in such more demanding reading-related tasks as SiT, leading to a higher mean duration. This understanding also offers a complementary perspective to explain why the overall fixation counts and regressions are higher in SiT than in normal reading for comprehension.

Carver (1990) systematically reviews relevant studies and concludes that reading consists of five processes, including scanning, skimming, rauding (i.e., normal reading), learning and memorising (pp. 12–22), which are utilised by readers to fulfil the goal of reading. As his focus is on reading speed, each process is associated with a standard (model) speed, represented in the format of Word per minute (Wpm). Here the *Word* denotes a standard-length word stretching six-character spaces (Carver 1977) – Wpm accounts for the word length variation, which intricately affects reading (Clifton et al. 2016). The representative Wpm and corresponding fixation duration per word for each process are reported by Carver (1990) as follows: scanning (600 Wpm & 100 ms), skimming (450 Wpm & 133 ms), rauding (300 Wpm & 200 ms), learning (200 Wpm & 300 ms), and memorising (138 Wpm & 433 ms). Worth mentioning is that (1) saccades generally shorten as the duration lengthens, and (2) the rauding process varies little across participants and tasks, while the other four processes can “vary more when executed by different individuals and...within individuals on different occasions or conditions” (Carver 1990: 22). The second point raised here is especially worth noting when comparing SiT reading in our study with Carver’s work. We anticipate seeing the same rauding process in our data as well, as

the results deriving from different tasks and participants in previous studies have converged on this front. On the other hand, we might see more variations in other processes or even processes that have not been found in previous research as a result of the tasks we use and such background variables as professional experience and training.

Other studies have similarly identified plural reading processes. Simola et al. (2008) report three processes in information search tasks, including (1) scanning (135 ms), manifested by saccades longer than rauding and fewer regressions, (2) normal reading (200 ms), with a saccade that extends around one word, and (3) decision making (175 ms), which has slightly more regressions than rauding but shares similar saccade length. Using a different task that requires participants to judge topic-text relevance, Olivier et al. (2022) have identified four processes: normal reading (304 wpm; 197 ms), fast reading (509 wpm; 118 ms), slow confirmation (263 wpm; 228 ms), and information search (183 wpm; 328 ms) (we added duration to make comparison easier). The two studies here have corroborated Carver's (1992) claim: Rauding remains similar across participants and tasks, while the other processes can take various forms according to the task in question and therefore fluctuate more.

Discerning what reading processes are involved in SiT complements the current SiT scholarship, as it offers a unique perspective to account for the fluctuations in fixation durations observed in different studies. Namely, fixation duration may reflect cognitive load, which seems to be the focus of the majority of the SiT studies, but it might not be the only reason. When examined together with saccade length and direction, shorter fixation duration with a longer outward regression could indicate a different reading approach, such as preparing to revert back to an earlier region of the text to locate the syntactic asymmetry between the source and the target languages to facilitate reorganising the sentence structure and delivering one's rendition following normal reading for comprehension (Lijewska et al. 2022). Moreover, as current research heavily centres on fixation duration and frequency, the inclusion of saccade can further help describe the features of SiT reading. Therefore, our study aims to understand SiT reading, drawing on the methodology and findings of the studies introduced in this subsection. Two indicators are considered to identify reading processes, including fixation duration and (outgoing) saccade, which also carries information about directionality. We calculate the number of crossed words (NCW) (Olivier et al. 2022) in place of absolute saccade length, which frequents in previous research. NCW is easier to interpret, as what absolute saccade length means varies based on the length of the fixated word(s). Cluster analysis is used to classify reading processes, inspired by Hyönä et al. (2002).

3 Methodology

This exploratory study presents the data from a larger research project investigating three tasks using an eye-tracker, namely silent reading (SR), reading aloud (RA), and SiT. Details are presented below.

3.1 Participants

The data of three groups of participants were analysed, including 17 experienced interpreters, 18 interpreting students, and 18 bilinguals without interpreter training. All participants considered themselves to be native Chinese speakers and deemed English as their first foreign language. The interpreters all had at least 150 days of professional experience, while the trainees all received equivalent postgraduate interpreter and SiT training. All participants met the language proficiency requirement (IELTS 6.5), had normal or corrected-to-normal vision, and signed the informed consent. The interpreters turned out to have significantly higher English proficiency ($M = 8.08$) than the trainees ($M = 7.58$) and the untrained bilinguals ($M = 7.27$), but the latter two were statistically equal; working memory size was comparable across groups (details see Ho 2017).

3.2 Materials and research design

Three 175-word English speech scripts were used. All were adapted from different speeches by the same speaker and offered an overview of the World Trade Organization (WTO). As Table 1 shows, the three scripts are considered similar, regarding not just the percentages of passive sentences, Flesch reading ease score, and Flesch-

Table 1: Features and difficulty ratings of the source speech scripts.

	Text 1	Text 2	Text 3
Word count	175	175	175
Passive sentences (in %)	11	12	10
Flesch reading ease	44.0	53.1	39.6
Flesch-Kincaid grade level	11.9	11.2	12.0
Difficulty rating (1–7) by untrained bilinguals in the original project	3	2.75	3.25
Difficulty rating (1–7) by interpreting students involved in a follow-up project	4	3.92	4.08

Kincaid grade level, but also the difficulty level rated by both the untrained bilinguals in the original project and the interpreting students recruited in a follow-up project.

Each participant had to conduct SR, RA, and L2-L1 SiT once, each using a different script. The order of the tasks and scripts were counter-balanced within and across groups. Being too long to fit into one computer screen ($1,024 \times 768$ pixels), each script was shown on four consecutive screens (named *trials*; up to six lines of text per trial). Sentences in every trial were complete, meaning the participants did not have to move to the next one to finish reading a sentence. Going backward to previous screens was not possible. The scripts were projected on a grey background in 22-point Courier. Eye movement data were sampled at 1,000 Hz using Eyelink 1000. Each participant was seated at around 70 cm from the monitor.

3.3 Procedure

The experiment began with an introduction to the research project and the self-paced tasks involved. The participants were advised to conduct the SiT task with an audience in mind, as if they were hired to interpret at a real conference (Setton and Dawrant 2016). They were also informed that they could not scroll back to previous trials once they moved on. Nine-point calibration then followed, and, if successful, the participants would engage with all three tasks in turn. Each task was preceded by a warm-up task of the same nature and followed by two comprehension questions to ensure genuine engagement. Eye-tracking accuracy was recalibrated in between tasks and trials when necessary.

3.4 Data analysis

The data from a total of 53 participants mentioned previously in Section 3.1 were included. For the purpose of the current study, fixations shorter than 80 ms and longer than 1200 ms were removed (Drieghe et al. 2008; Ma et al. 2021; White 2008) and 6.8 % of the data were hence deleted. Two features of each fixation were then chosen for cluster analysis: fixation duration and NCW, which represents the length of outward saccade plus its direction. Around 3 % of the data without NCW were further deleted. Finally, a total of 57,805 fixations underwent cluster analysis using the Partitioning Around Medoids (PAM) algorithm (Kaufman and Rousseeuw 1990). We analysed the full dataset together instead of examining each task separately, although we acknowledge each task may require distinct processes. The rationale is that the fundamental processes, such as decoding words and integrating meaning, are still shared across tasks (see Section 2.2 on overlapping processes between

studies on various reading-related tasks). SiT studies investigating reading passes have also shown that reading is similar between normal reading for comprehension and SiT in the first pass, substantiating the claim of shared processes to some extent (Ho et al. 2020; Ho 2021; Lijewska et al. 2022), hence our decision of one single cluster analysis. Another major reason relates to how we used cluster analysis, which would single out unique processes if they were indeed peculiar to a certain task, as we did not pre-determine how many clusters our data should have but were rather informed by PAM (more details see below).

Clustering means partitioning a dataset into “clusters”, with data points within the same cluster sharing more similarity than those in different ones. The idea is to identify clusters that are as distinctive as possible from each other (Schubert and Rousseeuw 2019). Prior to any analysis, the dataset was first examined using the Hopkins statistic (H), which “is used to assess the clustering tendency of a dataset by measuring the probability that a given data set is generated by a uniform data distribution. In other words, it tests the spatial randomness of the data” (Kassambara 2017: 123). An H -value close to 0.5 means the data set is uniformly distributed, hence no possibility of finding genuine clusters. The smaller the H -value, the more possible the dataset can be meaningfully clustered. Our data turned out to be highly clusterable ($H = 0.018$).

PAM was used in our study as a widely used algorithm and a robust alternative to k-means clustering. For k-means, a cluster centre is the mean value of all data points in the said cluster, whereas for PAM the centre (called *medoid*) of a cluster is the one presenting minimal dissimilarity to all the rest data points, making it “the most centrally located point in the cluster” (Kassambara 2017: 48), hence less sensitive to outliers. To know how many clusters will best partition the dataset and achieve best compactness for each cluster (i.e., minimal within-cluster difference) while maximising the average distance between clusters (Brock et al. 2008), silhouette width index (Rousseeuw 1987) was utilised. The silhouette width for each fixation can range from -1 to $+1$. The higher the value, the better, whereas anything below 0 indicates misclassification (ibid.).

Based on previous studies on identifying reading processes using fixation duration and saccade length (e.g., Carver 1990; Simola et al. 2008), we anticipated the possibility to find three to seven different reading processes and therefore examined the average silhouette width of each of these possibilities. We then picked the one with the highest average silhouette width and used the REMOS algorithm (version 2) (Lengyel et al. 2021) to optimise cluster allocation and ensured all data points ended with positive silhouette widths.

Each fixation in our data was assigned to only one cluster. The mean frequency of each cluster by task and by group was then calculated and then log-transformed before conducting between-subject and within-subject analysis using repeated measures ANOVA.

4 Results

This section consists of three parts. We first present the clustering results across tasks and groups and describe the features of clusters and compare them to the reading processes identified in the literature. We then report the overall statistical results between groups and tasks in the second section. Due to the limitation of space, within-subject analysis will be covered in a separate article and here the results are meant to provide a backdrop against which we compare group behaviours. We therefore focus on group comparisons in the final section.

4.1 Cluster analysis

The analysis resulted in five distinct clusters (i.e., reading processes). Table 2 presents the mean, median, and standard deviation (SD) of fixation durations and NCWs for each cluster. The number of clusters found in this study intuitively corroborates Carver’s (1990) findings, but the features of some clusters are worlds apart.

Fixation count shows that Cluster 1 & 2 are the majority, accounting for around 36 % and 45 % of all fixations respectively, while the frequencies of the other three clusters range from infrequent to rare. Cluster 1 has an MFD of 150.4 ms (SD = 33.21 ms; Mdn = 156 ms), akin to what Carver (1990) defines as skimming. On the other hand, Cluster 2 has an MFD of 260 ms (SD = 39.31 ms; Mdn = 255 ms). This seems to sit between the process of rauding (200 ms) and learning (300 ms) stated by Carver (1990), whose results about rauding have further been corroborated by Simola et al. (2008) and Olivier et al. (2022) even in tasks of different nature.

Starting from Cluster 3 onward, appearance is much more infrequent, ranging from around 15.3 % for Cluster 3 ($M = 423.97$ ms; $SD = 62.44$ ms; $Mdn = 409$ ms), 2.8 % for Cluster

Table 2: Descriptive statistics of fixation durations and NCWs for each cluster.

Cluster	Fixation count	Fixation duration			Number of crossed words (NCW)		
		Mean	Median	SD	Mean	Median	SD
1	20,802	150.42	156	33.21	0.5	1	3.52
2	26,067	260.22	255	39.31	0.68	1	2.81
3	8,842	423.97	409	62.44	0.62	1	3.21
4	1,608	735.85	689	142.18	0.66	1	3.98
5	486	264.24	222.5	158.58	−31.83	−30	10.99

4 ($M = 735.85$ ms; $SD = 142.18$ ms; $Mdn = 689$ ms), to 0.8 % for Cluster 5 ($M = 264.24$ ms; $SD = 158.58$ ms; $Mdn = 222.5$ ms). These three types of fixations are what really differentiate our findings from previous studies. While Cluster 3 can still claim to be similar, in terms of how demanding it is, to the memorising process proposed by Carver (1990), Cluster 4 is nowhere close to any of the categories presented in the literature about reading processes. Cluster 5 seems to be reflecting a categorically different process when compared with Cluster 2, with the SDs of the two clusters being wide apart.

The uniqueness of Cluster 5 is further substantiated by Figure 1 below. It is expected to see the SD widening from Cluster 1 to Cluster 4, with the increasing variability reflecting the more demanding nature of the process in question. However, Cluster 5 shows the utmost level of fluctuation, suggesting that this type of fixation indicates a different cognitive process that leads to wide variability across tasks and participants.

A look at NCW can shed more light on what Cluster 5 represents. For NCWs, all except Cluster 5 have a median of 1 NCW, which means the fixation jumps onto the immediately following word for half of the time. Cluster 5 stands out on this indicator with a mean NCW of -31.83 ($Mdn = -30$), and a much wider SD of 10.99. We further checked the range and realised it is exclusively a “regressive” cluster, with the shortest regression shooting across 17 words to earlier parts of the text.

After examining the MFD and NCW of each cluster, we argue that Cluster 1 reflects the skimming process, while Cluster 2, the most common type of fixation, is probably rauding (in Carver’s terms) – or normal reading for comprehension – albeit the MFD is slightly higher than what Carver (1990) reports (but in Rayner 2009 the reported range of

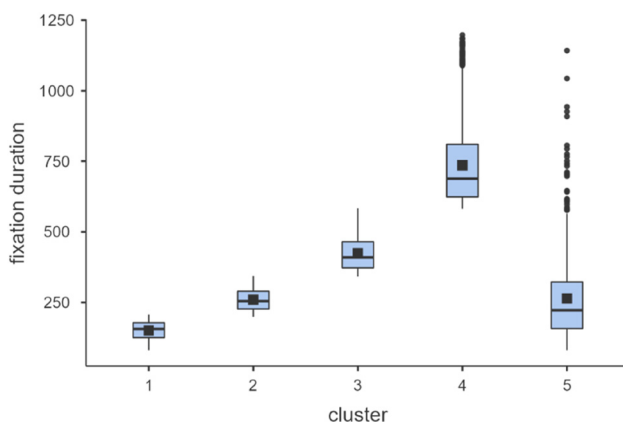


Figure 1: Boxplots showing the mean (square), median (horizontal line), and the distribution of each cluster.

SR overlaps to a large extent with our findings; also see Shreve et al. 2010, which uses bilingual reading and shows our findings are comparable). Cluster 3 & 4 could represent similar processes that are equally or more demanding than the memorising process mentioned by Carver (1990). Here, we would hypothesise that, due to the formality of the source scripts and the context in which these speeches take place in real life, the two clusters could be lumped together to indicate problems experienced by the participants, with Cluster 4 reflecting problem-triggers of a greater magnitude – semantically or syntactically. We therefore label these two clusters as “problem-solving” processes. Lastly, Cluster 5 could be “anchoring” fixations. That is, this type of fixation signals long regressions for the reprocessing or integration of syntactically dense sentences, which could stretch across four lines of text on the screen in this study.

4.2 Differences and similarities in the reading processes across groups and tasks

Table 3 presents the mean frequency distributions of the five clusters for each group across the tasks. This information, together with Figure 2 on the distribution of clusters within each group, mainly helps us tackle the first research question – what reading processes are involved in SiT and how they differ in SR and RA respectively – but also partly addresses the second question on how different groups vary in the combination of processes when tackling different tasks. Table 3 shows that the total frequency of fixations in SR and RA sits in the range around 200–300, but SiT is on a different level: around 400 to 800. Using SR as the baseline, the between-group gaps in frequency seem to be the smallest in RA and dramatically widen in SiT, though mainly between untrained bilinguals and the rest two groups. Another observation is that untrained bilinguals almost consistently had the highest mean fixation count for all clusters in all tasks, followed by interpreting trainees, and then experienced interpreters. There were only few exceptions, including Cluster 4 in SR and RA – where trainees had the highest number – and Cluster 5 in SiT, which experienced interpreters used almost equally often as trainees. As all the interpreters except one received the same training as the trainees,² a clear message is that training here does exert an obvious influence on the overall reading behaviour, leading to much fewer fixations in SiT.

Figure 2 visualises the use of various reading processes across groups in the three tasks. The percentage is presented to focus on the relative proportion of fixation counts for the five clusters within each group and each task.

² All experienced interpreters (except one) and trainees had been enrolled in the same two postgraduate programmes in Taiwan, which have frequent exchanges. The only exception received training from a renowned UK Interpreting programme at postgraduate level.

Table 3: Mean fixation counts of each cluster across groups and tasks.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Total
SR^a						
BIL ^b	121	148	35	3.4	2.8	310.2
INT ^b	101	126	33	3.7	2.4	266.1
PRO ^b	87.5	105	20.9	1.5	2.2	217.1
RA^a						
BIL	103	137	53.9	6.4	2.9	303.2
INT	79.3	122	53.1	8.9	2.4	265.7
PRO	76.1	115	44.4	6.1	2.1	243.7
SiT^a						
BIL	277	346	128	28.6	6.3	785.9
INT	174	193	72.6	17.2	3	459.8
PRO	154	179	56.9	14.8	3.2	407.9

^aSR, silent reading; RA, reading aloud; SiT, sight interpreting/translation. ^bBIL, untrained bilinguals; INT, interpreting trainees (novices); PRO, experienced interpreters.

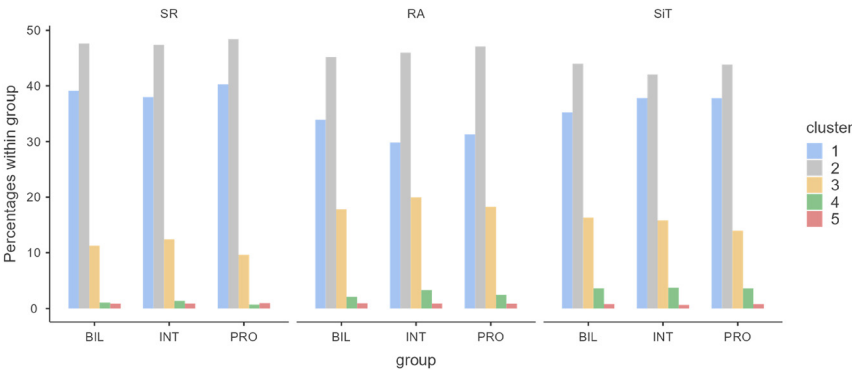


Figure 2: Fixation distributions (in percentage) of the five clusters for each group across tasks.

As Figure 2 illustrates, the pattern of distribution for all tasks is largely similar. Cluster 2 (rauding) is the most common type, followed by Cluster 1 (skimming), Cluster 3 (problem-solving), Cluster 4 (effortful problem-solving), and then Cluster 5 (anchoring). The last two clusters show minimal presence.

Using SR as the baseline, we can see how percentages of processes vary when there is an additional subtask of articulation in RA and when additional subtasks of interlingual reformulation and articulation are required in SiT. Across all tasks, the gap between Cluster 1 & 2 in RA seems to be the largest of all tasks; on the contrary, the same gap in SiT appears to be the smallest. In addition, Cluster 3 in RA accounts

for the largest proportion, followed by SiT and then SR. In terms of Cluster 4, there is an obvious uptick in RA and SiT when compared with SR.

We used repeated measures ANOVA to examine the effects of group and task on the frequency of each cluster. Omega squared was chosen to report effect size and was manually calculated following the instructions of Mellinger and Hanson (2017). Table 4 demonstrates three major phenomena: (1) Cluster 1 & 5 report effects of group and task, without interaction; (2) Cluster 2 & 3 show effects of group and task with interaction; (3) Cluster 4 only has a significant effect of task with interaction between group and task.

The variable of group has an significant impact on the frequency of almost all clusters: Cluster 1, $F(2, 50) = 8.16, p = <0.001, \omega^2 = 0.213$; Cluster 2, $F(2, 50) = 12.5, p = <0.001, \omega^2 = 0.303$; Cluster 3, $F(2, 50) = 8.65, p = <0.001, \omega^2 = 0.224$; Cluster 5, $F(2, 50) = 3.53, p = 0.037, \omega^2 = 0.087$. Cluster 4 is the only non-significant cluster, $F(2, 50) = 2.09, p = 0.134, \omega^2 = 0.04$.

On the other hand, all clusters show statistically significant effects of tasks. That is, the task conducted does affect the use of reading processes differently, including Cluster 1, $F(2, 100) = 113.05, p = <0.001, \omega^2 = 0.809$, Cluster 2, $F(2, 100) = 95.98, p = <0.001, \omega^2 = 0.782$, Cluster 3, $F(2, 100) = 172.21, p = <0.001, \omega^2 = 0.866$, Cluster 4, $F(2, 100) = 95.63, p = <0.001, \omega^2 = 0.781$, and Cluster 5, $F(2, 100) = 8.64, p = <0.001, \omega^2 = 0.224$. The effect size is surprisingly large throughout, except for Cluster 5. On top of this, how reading processes are used differently in each task is significantly related to some extent with which group the participant belongs to for Cluster 2, $F(4, 100) = 6.54, p < 0.001, \omega^2 = 0.295$, Cluster 3, $F(4, 100) = 7.73, p < 0.001, \omega^2 = 0.337$, and Cluster 4, $F(4, 100) = 2.87, p = 0.027, \omega^2 = 0.124$.

Table 4: Analysis of the differences between cluster frequencies using repeated measures ANOVA.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Group	<i>F</i>	8.16	12.5	8.65	2.09	3.53
	<i>p</i>	<0.001	<0.001	<0.001	0.134	0.037
	ω^2	0.213	0.303	0.224	0.04	0.087
Task	<i>F</i>	113.05	95.98	172.21	95.63	8.64
	<i>p</i>	<0.001	<0.001	<0.001	<0.001	<0.001
	ω^2	0.809	0.782	0.866	0.781	0.224
Group*Task	<i>F</i>	1.9	6.54	7.73	2.87	1.91
	<i>p</i>	0.116	<0.001	<0.001	0.027	0.114
	ω^2	0.0636	0.295	0.337	0.124	0.064

Table 5: Group post-hoc comparisons for Cluster 2.

Task	Group		Task	Group	<i>t</i>	<i>p</i> _{bonferroni}
SR ^a	BIL ^b	–	SR	INT ^b	1.712	1
SR	BIL	–	SR	PRO ^b	3.261	0.072
SR	INT	–	SR	PRO	1.574	1
RA ^a	BIL	–	RA	INT	1.625	1
RA	BIL	–	RA	PRO	1.927	1
RA	INT	–	RA	PRO	0.325	1
SiT ^a	BIL	–	SiT	INT	4.996	< 0.001
SiT	BIL	–	SiT	PRO	5.169	< 0.001
SiT	INT	–	SiT	PRO	0.244	1

^aSR, silent reading; RA, reading aloud; SiT, sight interpreting/translation. ^bBIL, untrained bilinguals; INT, interpreting trainees (novices); PRO, experienced interpreters.

4.3 Reader differences come from training, not professional experience

For this subsection, we delve into the post-hoc analysis (with Bonferroni correction) of the between-subject effects. There is no interaction effect for Cluster 1, which means the difference between groups observed does not vary across tasks. The untrained bilinguals had significantly more Cluster 1 fixations (skimming) than the trainees, $t(50) = 3.019$, $p = 0.012$, and the experienced interpreters, $t(50) = 3.823$, $p = 0.001$, respectively. The latter two groups were statistically identical, $p = 1$.

Cluster 2 presents a more complicated relationship between task and group, as shown in Table 5. It turns out that there was no between-group difference in SR or RA. By contrast, in SiT the untrained bilinguals used significantly more Cluster 2 fixations (rauding) than trainees, $t(50) = 4.996$, $p = <0.001$, and experienced interpreters, $t(50) = 5.169$, $p = <0.001$. Similar to Cluster 1, there was no difference between the latter two groups in all tasks, $p = 1$ throughout.

For Cluster 3, the detailed comparisons are included in Table 6. The main group effect has three contributors. The untrained bilinguals resorted to problem-solving processes significantly more than experienced interpreters in SR, $t(50) = 3.3899$, $p = 0.049$. More importantly, in SiT, the cluster frequency of the bilinguals was much higher than that of the trainees, $t(50) = 4.057$, $p = 0.006$, and the interpreters, $t(50) = 5.8035$, $p = <0.001$.

Table 6: Group post-hoc comparisons for Cluster 3.

Task	Group		Task	Group	<i>t</i>	<i>p</i> _{bonferroni}
SR ^a	BIL ^b	–	SR	INT ^b	0.6927	1
SR	BIL	–	SR	PRO ^b	3.3899	0.049
SR	INT	–	SR	PRO	2.7072	0.333
RA ^a	BIL	–	RA	INT	0.4928	1
RA	BIL	–	RA	PRO	1.3605	1
RA	INT	–	RA	PRO	0.8748	1
SiT ^a	BIL	–	SiT	INT	4.057	0.006
SiT	BIL	–	SiT	PRO	5.8035	< 0.001
SiT	INT	–	SiT	PRO	1.8048	1

^aSR, silent reading; RA, reading aloud; SiT, sight interpreting/translation. ^bBIL, untrained bilinguals; INT, interpreting trainees (novices); PRO, experienced interpreters.

The disparities between groups are most noticeable from Cluster 1 to Cluster 3, and there is no group effect for Cluster 4. As for Cluster 5, although the main effect was significant, post-hoc analysis shows no difference between any two groups ($p = 0.067$ for BIL vs. INT, $p = 0.091$ for BIL vs. PRO, and $p = 1$ for INT vs. PRO).

5 Discussion

Two major strands of findings are used to address the two research questions of this study. Firstly, cluster analysis was conducted to address the first research question, which inquired about the reading processes in SiT and how task specificity affects their use when compared with SR and RA. Five distinct reading processes were found, coinciding with Carver’s (1990) conclusion after reviewing studies on various reading tasks.

However, the match is only partial because of the mismatch in the MFD for some clusters and the fact that the second indicator we used was slightly different, i.e., NCW instead of absolute saccade length. Cluster 1 & 2 exhibit close approximation to the types of fixations documented in previous studies on reading (in addition to Carver 1990, see also Olivier et al. 2022; Simola et al. 2008). Cluster 1 much resembles skimming in terms of fixation duration. Considering that the function of skimming is to capture the most important ideas or the gist of a text (Sulaeman et al. 2021), especially under time pressure (Duggan and Payne 2009), it makes sense to see this reading process in our tasks as they all require, if not more than, comprehension. That aside, SiT creates a perfect environment that prompts skimming. To successfully perform SiT, practitioners have long been advised to read faster and ahead (Agri-foglio 2004; Chen 2015; Lee 2012) when reading, memorisation, reformulation, and

articulation all have to be coordinated (Gile 2009) and executed under time pressure (Mikkelsen 1994) and in a seemingly easy and smooth manner (Angelelli 1999).

Cluster 2, on the other hand, is arguably the rauding process. First and simply, it is the most frequent process used in all three tasks, and it is difficult to fathom another process being used more than the normal reading process that aims at comprehension. Aside from the match (within reasonable range) with the relevant literature regarding fixation duration, a median of 1 for NCW and a mean closer to 1 plus a much smaller SD than other clusters show the tendency of Cluster 2 to proceed in general on a word-to-word basis, which is expected to be preferable or necessary for scripts used in our study. In fact, the slightly higher MFD might be explained by the same reason: The marginally slower rauding rate could be a countermeasure to ensure maximal efficiency of comprehension (Carver 1990) in a high-stake context.

Cluster 3 fits within the existing framework about reading processes between learning and memorising (and approximating the latter), the two more demanding processes. Yet, a glance at the description of what they could entail (see Carver 1990), we find that both could adequately describe Cluster 3, including but not limited to encountering infrequent terms, meaning integration for syntactically complex sentences, memorising important ideas, or even retaining terms or meaning segments in memory to later structure idiomatic and syntactically correct rendition in SiT. Cluster 4, with all the similarities shared with Cluster 3, could well represent the same process. When also considering the (in)frequency of these two processes, we deem “problem-solving” processes a suitable label for these two categories. Bear in mind that *problem* here “does not necessarily mean a difficult or troubling situation... Problem solving is any activity that involves original thinking to develop a solution, solve a dilemma, or create a product” (Kapp 2012: 144).

Lastly, Cluster 5 appears to be unique in that it represents not just regressions, but extreme regressions. Judging from the length it frequently travelled, these fixations likely happened when there were long, complex, and compound sentences in the source scripts, which were commonplace for speeches given in formal settings on the global stage, such as the WTO meeting simulated in our study. The adjectival or adverbial parts of these long and challenging sentences are structured inversely in terms of their relative positions to what they modify across the levels of phrases, clauses, and sentences between Chinese and English (Chen 2006; Yang et al. 2010), hence an even higher likelihood to prompt regressions (also see Ma 2021; Ma et al. 2021; Ma et al. 2022).

After distilling five different clusters of fixations from our dataset, we compared the combination of these processes in different tasks. The total fixations of SiT, as shown in Table 3, have greatly increased for all groups, compared to SR and RA (see also Jakobsen and Jensen 2008; Wang and He 2018; but see Alves et al. 2011 for opposite findings). However, the extent of increase delineates a more nuanced

picture. In SR, the difference can probably be attributed to language proficiency. In RA, the interpreter group edges much closer to the other two groups – which might have originated from the requirement of articulation. This additional subtask could have slowed down the professionals (relatively to their own processing in SR), as coordination of reading and articulation leaves little capacity to such subtasks (Inhoff and Radach 2014) as tackling complex meaning or syntactic units. On the other hand, SiT clearly sets the trained and untrained participants apart, both in the fixation frequency of each cluster and the total fixation count. Interestingly, Figure 2 shows that, percentage-wise, the overall reading patterns in all tasks remain similar between groups (and even between tasks): Cluster 2 is the most frequent, followed by Cluster 1, 3, 4, and 5 in order.

As mentioned, all groups had statistically similar working memory sizes. The interpreters and trainees had almost identical training, so the difference between the two groups comes down to professional experience and language proficiency. Meanwhile, the only difference between trainees and untrained bilinguals is training. Therefore, by asking the second research question – how participants in different groups read – we shall see the effects of training (trained vs. untrained), and professional experience plus language proficiency (experienced interpreters vs. the other two groups). The results show significant differences by group regarding the frequency of each cluster. There are clear effects of training through Cluster 1–3. Considering the extremely rare occurrence of Cluster 4 (effortful problem-solving) and Cluster 5 (anchoring), similar behaviour is expected in tackling formal and complex speech scripts. Overall, training seems to have helped reduce how many times participants read the scripts. The repeated emphasis in interpreter training on deeper and swift analysis (Gile 2009; Liu 2020; Setton and Dawrant 2016) may have led to the trained participants being more effective and decisive, with language proficiency and experience of dealing with scripts of similar nature further widening the gap between the bilinguals and the experienced interpreters. While skimming was consistently used to different extents for all tasks between groups, the reading process was statistically identical across groups, except in SiT. Similarly, the difference in problem-solving behaviour (Cluster 3) was almost exclusive in SiT, except that language proficiency might have given the professionals some advantage over the bilinguals. The findings tell us that all participants are practically proficient readers, with almost no difference in baseline reading tasks. Hence, the effects of training are most potent in SiT, with frequencies almost halved consistently across all reading processes (while significantly improving SiT performance in all aspects, see Ho et al. 2020).

Interestingly, the impact of professional experience (and/or language proficiency) is almost invisible, as there is no difference between the trainees and the interpreters in cluster-based frequencies as local indicators (see also Ho 2021 about

word-based analysis). As the two groups have received the same training at post-graduate level, this finding has a three-fold implication: firstly, training for one semester to a year is sufficient for the trainees to adopt similar reading approaches to the experienced interpreters with at least 150 days of experience (in the case of the Taiwanese market this could mean around 5–7 years after a graduate enters the market); secondly, training is efficient and effective as it helps the trainees achieve a similar level of mastery in tackling complicated speech scripts; thirdly, one semester of training is quick and effective, but a second semester-long training might not add significant changes (as some trainees and professionals received only one-semester SiT training, also see Chmiel and Mazur 2013; cf. Fang et al. 2023). These observations also corroborate that of Ho (2021), in which the trainees were surprisingly similar to the experienced interpreters in many of their SiT and reading behaviour and were capable of achieving an equivalent level of delivery (called *style* in that study). It is possible that the principles, skills, and strategies for successfully conducting SiT can be taught (or at least made aware) in one semester. Following from this, as reported in Ho (2021), the major difference between the two groups will be the actual repeated use of the skills that builds up the “reflex”, which takes time, plus the fundamental language proficiency. The combined effect of both factors probably led to the experienced interpreters’ higher accuracy (also see McDonald and Carpenter 1981), a quicker pace after oral rendition began, and fewer and shorter observable pauses, which all reach significance statistically (Ho 2021).

6 Conclusions

Inspired by reading research, this study set out to examine SiT reading processes, which have not yet been explored in relevant research. Understanding reading processes provides us with a new perspective to better understand the nature of SiT and the cognitive processes therein, as reading processes are defined not by fixation duration alone, which is the sole focus of the majority of relevant studies, but in conjunction with saccade length and direction. The results can help us move away from the sole focus on cognitive load, indexed only by fixation duration; rather, different combinations of fixation duration and saccade length (and direction) could signal various reading processes or approaches, thereby offering possible alternative explanations for the “partial” phenomena we observe, e.g., when fixation duration increases or decreases.

We adopted cluster analysis to examine the reading processes used by participants during SR, RA, and SiT. Five reading processes have been identified in the current dataset, consisting of untrained bilinguals, interpreting trainees, and experienced interpreters. The most common reading process is Cluster 2 (rauding) – normal reading process for comprehension. The second frequent process is Cluster 1, which

largely resembles skimming. The two categories together account for around 81.1 % of all fixations. Cluster 3 & 4 are both problem-solving processes, with the former accounting for 15.3 % and the (more effortful) latter dropping to 2.8 %. Cluster 5 (0.8 %) is rare and distinct in that it exclusively precedes extremely long regressions for further processing, hence named “anchoring” process.

The two research questions in our study intend to understand how the combination of reading processes varies according to the nature of the task at hand and how the three groups of participants approach reading differently in each task. Our main results show that the group-based combination of processes is largely similar across tasks, proportional to the overall percentage of each cluster.

Lastly, how our three groups of participants used reading processes differently mainly show the effects of training. While the reading behaviour is statistically the same between experienced interpreters and trainees across all clusters, there is a clear divide between the trained participants and the untrained bilinguals, with the latter having significantly more fixations for skimming, rauding, and typical problem-solving. More specifically, the impact is visible almost exclusively in SiT. This shows that the three groups are equally proficient in tackling the source scripts in SR and RA, proving that the basic reading abilities are similar. Considering findings from previous research on top of the discussion above, training seems to make reading more effective and decisive in SiT, so fewer fixations are needed.

A major limitation of our study is that, while we believe SiT is a situated communicative activity, the research design was not able to capture the dynamic interactions with audience, as we were limited by the desktop eyetracker (for a good example, see Chmiel and Lijewska 2019 and how they tried to partially mitigate the drawback). Future studies are advised to use wearable eyetrackers to capture all behaviours, including interactions with audience and other environmental elements, which will in turn improve the ecological validity.

In addition, our findings are only exploratory and by no means final. Many other factors could still affect the results, including participant background factors, such as language pair, language proficiency, and level/type of training, and experiment conditions, e.g., whether real audience are present for communication purposes. We therefore encourage follow-up studies to verify the findings and even compare across language pairs to identify universal and language-specific reading behaviours in SiT.

Due to the limitation of space, we only presented the overall frequencies and percentages of reading processes. This may be another limitation when trying to explain the dynamic SiT reading behaviours, as different reading passes might present utterly distinct or even opposite results. Therefore, follow-up investigation of cluster distribution in each reading pass is needed, as in this study, a fair share of

fixations in several clusters, such as Cluster 1 & 2 certainly entail refixations and/or (shorter) regressions. Refining the results based on reading passes will certainly help shine a light on the cognitive processes at work, especially in SiT, where reformulation in a different language is an integral part of the task. Pinpointing where Cluster 3, 4, and 5 fixations land in the text will also further unveil the complex reading behaviour in SiT.

References

- Agrifoglio, Marjorie. 2004. Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting* 6(1). 43–67.
- Alves, Fabio, Adriana Pagano & da Silva, Igor. 2011. Towards an investigation of reading modalities in/for translation: An exploratory study using eye-tracking data. In Sharon O'Brien (ed.), *Cognitive explorations of translation*, 175–196. London: Continuum.
- Angelelli, Claudia V. 1999. The role of reading in sight translation: Implications for teaching. *The ATA Chronicle* 28(5). 27–30.
- Brock, Guy, Vasyl Pihur, Susmita Datta & Somnath Datta. 2008. cIValid: An R package for cluster validation. *Journal of Statistical Software* 25(4). 1–22.
- Carver, Ronald P. 1977. Toward a theory of reading comprehension and reading. *Reading Research Quarterly* 13(1). 8–63.
- Carver, Ronald P. 1990. *Reading rate: A review of research and theory*. San Diego, CA: Academic Press.
- Carver, Ronald P. 1992. Reading rate: Theory, research, and practical implications. *Journal of Reading* 36(2). 84–95.
- Čeňková, Ivana. 2015. Sight translation/interpreting. In Franz Pöchhacker (ed.), *Routledge encyclopedia of interpreting studies*, 374–375. London: Routledge.
- Chen, Jyun-gwang. 2006. *Contrastive research & crosslinguistic influence*. Taipei: Crane.
- Chen, Wallace. 2015. Sight translation. In Holly Mikkelsen & Renée Jourdenais (eds.), *The routledge handbook of interpreting*, 144–153. Abingdon: Routledge.
- Chmiel, Agnieszka & Agnieszka Lijewska. 2019. Syntactic processing in sight translation by professional and trainee interpreters: Professionals are more time-efficient while trainees view the source text less. *Target* 31(3). 378–397.
- Chmiel, Agnieszka & Iwona Mazur. 2013. Eye tracking sight translation performed by trainee interpreters. In Catherine Way, Sonia Vandepitte, Reine Meylaerts & Magdalena Bartłomiejczyk (eds.), *Tracks and treks in translation studies*, 189–205. Amsterdam: John Benjamins.
- Clifton, Charles, Fernanda Ferreira, John M. Henderson, Albrecht W. Inhoff, Simon P. Livsedge, Erik D. Reichle & Elizabeth R. Schotter. 2016. Eye movements in reading and information processing: Keith Rayner's 40year legacy. *Journal of Memory and Language* 86. 1–19.
- Drieghe, Denis, Alexander Pollatsek, Adrian Staub & Keith Rayner. 2008. The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(6). 1552–1560.
- Duggan, Geoffrey B. & Stephen J. Payne. 2009. Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied* 15(3). 228–242.
- Fang, Jing, Xiaomin Zhang & Haidee Kotze. 2023. The effects of training on reading behaviour and performance in sight translation: A longitudinal study using eye-tracking. *Perspectives* 31(4). 655–671.

- Gile, Daniel. 2009. *Basic concepts and models for interpreter and translator training*, Revised edition. Amsterdam: John Benjamins.
- Gorszczyńska, Paula. 2020. Disfluencies in sight translation vis-à-vis dominating text function: A pilot study based on English-Polish sight translation performed by professional interpreters. *Beyond Philology An International Journal of Linguistics, Literary Studies and English Language Teaching* 17(4). 95–130.
- He, Yan & Jiayi Wang. 2021. Eye tracking uncertainty management in sight translation: Differences between professional and novice interpreters. In Ricardo Muñoz Martin, Sanjun Sun & Defeng Li (eds.), *Advances in cognitive translation studies*, 181–200. Singapore: Springer Nature Singapore.
- Ho, Chen-En. 2017. *An integrated eye-tracking study into the cognitive process of English-Chinese sight translation: Impacts of training and experience*. Taipei: National Taiwan Normal University dissertation. Available at: <https://api.lib.ntnu.edu.tw:8443/server/api/core/bitstreams/5b6789ad-d1b8-4499-b83e-fe72bd56ce96/content>.
- Ho, Chen-En. 2021. What does professional experience have to offer? An eyetracking study of sight interpreting/translation behaviour. *Translation, Cognition & Behavior* 4(1). 47–73.
- Ho, Chen-En. 2022. Sight interpreting/translation. In *The Encyclopedia of translation & interpreting*. Iberian Association for Translation and Interpreting Studies (AIETI).
- Ho, Chen-En, Tze-Wei Chen & Jie-Li Tsai. 2020. How does training shape English-Chinese sight translation behaviour? An eyetracking study. *Translation, Cognition & Behavior* 3(1). 1–24.
- Ho, Chen-En & Yao Xiao. 2024. Reading processes in English-Chinese sight interpreting/translation tasks. In Riccardo Moratto & Cheng Zhan (eds.), *The routledge handbook of Chinese interpreting*, 186–206. London: Routledge.
- Hu, Ting, Xinyu Wang & Haiming Xu. 2022. Eye-tracking in interpreting studies: A review of four decades of empirical studies. *Frontiers in Psychology* 13. 872247.
- Hyönä, Jukka, Robert F. Lorch, Jr. & Johanna K. Kaakinen. 2002. Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology* 94(1). 44–55.
- Inhoff, Albrecht W. & Ralph Radach. 2014. Parafoveal preview benefits during silent and oral reading: Testing the parafoveal information extraction hypothesis. *Visual Cognition* 22(3–4). 354–376.
- Jakobsen, Arnt Lykke & Kristian T. H. Jensen. 2008. Eye movement behaviour across four different types of reading task. In Susanne Göpferich, Arnt Lykke Jakobsen & Inger M. Mees (eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing*, 103–124. Frederiksberg: Samfundslitteratur.
- Jiménez Ivars, Amparo. 2008. Sight translation and written translation. A comparative analysis of causes of problems, strategies and translation errors within the PACTE translation competence model. *FORUM* 6(2). 79–104.
- Kapp, Karl M. 2012. *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco: Pfeiffer & Company.
- Kassambara, Alboukadel. 2017. *Practical guide to cluster analysis in R: Unsupervised machine learning*. STHDA.
- Kaufman, Leonard & Peter J. Rousseeuw. 1990. *Finding groups in data: An introduction to cluster analysis*. New Jersey: John Wiley & Sons.
- Lee, Jieun. 2012. What skills do student interpreters need to learn in sight translation training? *Meta* 57(3). 694–714.
- Lengyel, Attila, David W. Roberts & Zoltán Botta-Dukát. 2021. Comparison of silhouette-based reallocation methods for vegetation classification. *Journal of Vegetation Science* 32(1). e12984.
- Li, Xiangdong. 2014. Sight translation as a topic in interpreting research: Progress, problems, and prospects. *Across Languages and Cultures* 15(1). 67–89.

- Li, Xiangdong. 2015. Designing a sight translation course for undergraduate T&I students: From context definition to course organization. *Spanish Journal of Applied Linguistics* 28(1). 169–198.
- Lijewska, Agnieszka, Agnieszka Chmiel & Albrecht W. Inhoff. 2022. Stages of sight translation: Evidence from eye movements. *Applied Psycholinguistics* 43(5). 997–1018.
- Liu, Jie. 2020. *Interpreter training in context: European and Chinese models reconsidered*. Singapore: Springer Singapore.
- Ma, Xingcheng. 2021. Coping with syntactic complexity in English–Chinese sight translation by translation and interpreting students. An eye-tracking investigation. *Across Languages and Cultures* 22(2). 192–213.
- Ma, Xingcheng, Dechao Li & Yu-Yin Hsu. 2021. Exploring the impact of word order asymmetry on cognitive load during Chinese–English sight translation: Evidence from eye-movement data. *Target* 33(1). 103–131.
- Ma, Xingcheng, Dechao Li, Jie-Li Tsai & Yu-Yin Hsu. 2022. An eye-tracking based investigation into reading behavior during Chinese–English sight translation: The effect of word order asymmetry. *Translation & Interpreting* 14(1). 66–83.
- Macizo, Pedro & M. Teresa Bajo. 2004. When translation makes the difference: Sentence processing in reading and translation. *Psicológica* 25(2). 181–205.
- Macizo, Pedro & M. Teresa Bajo. 2006. Reading for repetition and reading for translation: Do they involve the same processes? *Cognition* 99(1). 1–34.
- McDonald, Janet L. & Patricia A. Carpenter. 1981. Simultaneous translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior* 20(2). 231–247.
- Mellinger, Christopher & Thomas Hanson. 2017. *Quantitative research methods in translation and interpreting studies*. New York: Routledge.
- Mikkelsen, Holly. 1994. Text analysis exercises for sight translation. In *Proceedings of the thirty-first annual conference of the American translators association*. Medford: NJ Vistas.
- Nilsen, Anne Birgitta & May-Britt Monsrud. 2015. Reading skills for sight translation in public sector services. *Translation & Interpreting* 7(3). 10–20.
- Olivier, Brice, Anne Guérin-Dugué & Jean-Baptiste Durand. 2022. Hidden semi-Markov models to segment reading phases from eye movements. *Journal of Eye Movement Research* 15(4). 5.
- Rayner, Keith. 2009. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* 62(8). 1457–1506.
- Rousseuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65.
- Ruiz, Jason Omar & Pedro Macizo. 2019. Lexical and syntactic target language interactions in translation. *Acta Psychologica* 199. 102924.
- Ruiz, C., N. Paredes, P. Macizo & M. T. Bajo. 2008. Activation of lexical and syntactic target language properties in translation. *Acta Psychologica* 128. 490–500.
- Schubert, Erich & Peter J. Rousseeuw. 2019. Faster k-Medoids clustering: Improving the PAM, CLARA, and CLARANS algorithms. In Giuseppe Amato, Claudio Gennaro, Vincent Oria & Miloš Radovanović (eds.), *Similarity search and applications*, 171–187. Cham: Springer International Publishing.
- Setton, Robin & Andrew Dawrant. 2016. *Conference interpreting. A complete course*. Amsterdam: John Benjamins.
- Shreve, Gregory M., Isabel Lacruz & Erik Angelone. 2010. Cognitive effort, syntactic disruption, and visual interference in a sight translation task. In Gregory M. Shreve & Erik Angelone (eds.), *Translation and cognition*, 63–84. Amsterdam: John Benjamins.
- Simola, Jaana, Jarkko Salojärvi & Ilpo Kojo. 2008. Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research* 9(4). 237–251.

- Sulaeman, Agus, Achmad Suherman, E. Enawar & Supyan Sori. 2021. Skimming reading techniques on the ability to identify intrinsic drama text elements. *Journal of English Language and Literature* 6(1). 25–38.
- Wang, Jia-Yi & He Yan. 2018. How effortful are interpreters in translation related reading Tasks? An eye-tracking study. *Journal of Literature and Art Studies* 8(10). 1497–1508.
- White, Sarah J. 2008. Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance* 34(1). 205–223.
- Xiao, Yao, Kristian Tangsgaard Hvelplund & Chen-En Ho. 2023. Wearable eye trackers: Methodological challenges, opportunities and perspectives for sight interpreting/translation. *Translation, Cognition & Behavior* 6(2). 164–186.
- Yang, Chin Lung, Charles A. Perfetti & Ying Liu. 2010. Sentence integration processes: An ERP study of Chinese sentence comprehension with relative clauses. *Brain and Language* 112(2). 85–100.

Bionotes

Chen-En Ho

Queen's University Belfast, Belfast, Northern Ireland

c.ho@qub.ac.uk

<https://orcid.org/0000-0001-7321-2029>

Dr. Chen-En Ho (Ted Ho) is a Senior Lecturer in Translation and Interpreting in the School of Arts, English and Languages, Queen's University Belfast. His research interests now focus on two major strands, including cognitive translation and interpreting studies, primarily investigating sight interpreting/translation using an eye-tracker, and T&I education, particularly on autonomous learning and using the concept of gamification and feedback to improve learning experience and outcome.

Jie-Li Tsai

National Chengchi University, Taipei, Taiwan

jltsai@nccu.edu.tw

<https://orcid.org/0000-0001-5413-4889>

Dr. Jie-Li Tsai is an Associate Professor at Department of Psychology of National Chengchi University. In his Eye Movement and Reading Laboratory, the research works are mainly on psycholinguistics issues of reading Chinese and eye movement control and language processing. The research topics include lexical and contextual processing, parafoveal processing, semantic ambiguity resolution, reading non-native languages, and neural mechanisms of natural reading.