Eye Will Tell You – How Gaze Aversion Reveals Trainee Interpreters' Cognitive Load during Consecutive Interpreting Active Listening and Notetaking

Wenbo Guo

Doctor of Philosophy

School of Modern Languages, Newcastle University, UK

September 2022

Abstract

This thesis investigates trainee interpreters' Gaze Aversion (GA) - an overt strategy for regulating high cognitive load - from notepads during Consecutive Interpreting (CI) active listening and notetaking. The general aim of this project is to reveal the intricacy of GA in the context of CI active listening and notetaking and introduce it as a cognitive load indicator to Interpreting Studies. Specifically, the project fulfils several aims: to investigate if interpreters would conduct GA during active listening and notetaking when experiencing high cognitive load; to explore textual features that could potentially contribute to GA; to test if GA would enhance interpreters' performance. This PhD includes two studies: a corpus study and an experiment. The corpus study involved coding thirty videos of trainee interpreters performing English-to-Chinese CI. The experiment was conducted to corroborate the findings from the corpus study and further explore the intricacy of GA in a controlled environment. A mixed-methods design involving the coding of GA, speech difficulty ratings, observation of features of the source text preceding GA, and priming was adopted for the investigation. Results indicate that GA was observed among participants, but the behaviour was subject to considerable individual differences. GA was associated with various preceding textual features, but complex syntactic structures and low-frequency words were potent GA inducers. Results also suggest that although GA reliably indicated trainee interpreters' high cognitive load, it did not enhance interpreting performance as hypothesised in previous studies. Instead, GA was associated with inferior renditions. The absence of performance-enhancing benefits could be explained by the possibility that trainee interpreters' cognitive load reached over a tipping point where GA would be futile in circumventing cognitive overload, suggesting that GA could indicate cognitive crisis.

Acknowledgement

This PhD project has been the wildest dream of my life. Looking back at this journey, I see many people who deserve my unswerving gratitude.

First, I would like to thank my main supervisor, Dr Michael Jin, for his willingness to accept me as his PhD student. I would like to also thank him for offering knowledge, help, and guidance at the moments when they were most needed. My gratitude to Dr Jin also extends to his constructive criticism and professionalism as a mentor. I would also like to take this opportunity to thank my second supervisor, Prof Gwyneth Doherty-Sneddon for always being patient, warm, and smiley. I also thank her for explaining cognitive science and statistics to me whenever I needed help in those regards. Thanks also go to her for being so accommodating with me and looking after me. For both my supervisors, I want to thank them for showing me lights when there was darkness.

I want to thank my mother, Xiurong Wang, who always believes in me and encourages me to step out of my comfort zone. This PhD coincided with some life turmoil for both of us, but she never tried to hold me back even though when she was entitled to do so. Without her support, I wouldn't have made this far.

To Reginald, thank you for being you and in my life.

I must also offer my gratitude to Mr Eric Liu for his long-term support and guidance since 2014. He not only has the wisest words but also cooks the best lamb ribs. I feel lucky to have had him as a teacher and friend.

Thanks go to Nicki for arranging the supervision meetings. Thanks go to Martyn for being the speaker. Also, thanks are sent to NUORS committee for sponsoring this project.

I would like to thank myself for not giving up.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

This thesis is one of the earliest studies investigating gaze aversion (GA) in Consecutive Interpreting (CI) active listening and notetaking. GA originally refers to the behaviour of looking away from an interlocutor's face when a person needs to think of an answer to a challenging question (Doherty-Sneddon and Phelps, 2005). The rationale behind the behaviour is that the human face contains rich information and that looking at a face could interfere with processing the question if the question is cognitively challenging, such as an arithmetic question. By looking away from someone's face, a person can disengage from the distraction and re-route cognitive resources from processing the face to processing the question. Thus, GA is considered a spontaneous and overt strategy for regulating and indicating high cognitive load. Based on the load-regulating benefits of GA, scholars suggested that GA could boost cognitive performance. If the human face is distracting due to its capability of signalling rich information and considering human beings tend to process what they see (Carpenter and Just, 1980), it will make one wonder if the rationale and benefits of GA can be applied in other fields.

From the perspective of CI, scholars have left little room for doubt that the task is cognitively demanding, for it requires an interpreter to give unmitigated attention in active listening while mobilising linguistic and non-linguistic knowledge from memory (Cutler and Clifton, 1999). Perhaps the only visible effort during GA is notetaking, an activity where an interpreter takes advantage of the encoding and storage function of notes. It would be warranted to argue that interpreters' notes contain rich information, which would lead one to the question – if looking at a face interferes with answering a difficult question because the face contains too much information, would looking at a notebook interfere with interpreters' comprehension of the source speech? If so, interpreters might avert their gaze from the notepad to channel more resources to the task.

This thesis is dedicated to exploring GA in CI active listening and notetaking.

1.1. Overall Research Purpose

In interpreting studies, it is not uncommon for researchers to use eye movements as cognitive load indicators. Fixations and saccades patterns have been found to be associated with the fluctuation of cognitive load (e.g., Duchowski, 2017; Su, 2020). However, GA has not yet been studied as a potential cognitive load indicator. From the perspective of cognitive psychology, GA has not yet been introduced to fields that are seemingly remote. Therefore, the first overall research purpose is to bring interpreting studies and GA together. If GA proves to be a reliable cognitive load indicator, it will bear significant implications for both trainee and professional interpreters.

It would be warranted to argue that both GA and CI are intricate. Although GA is commonly seen, the detailed reasons behind the behaviour, such as what element would trigger high cognitive load, are not fully explored. In CI, although pioneers have made substantial contributions to understanding the task from the perspective of cognition, the dearth in researching CI active listening and notetaking is hard to ignore. Therefore, another overall research purpose is to gain a better understanding of CI via GA.

1.2. Research Methodology

Exploring GA in the sphere of CI active listening and notetaking, in this thesis, is achieved via an Initial Corpus Study and an Experimental Study. As previously mentioned, this thesis constitutes the earliest effort to investigate GA in CI. Albeit being ground-breaking, the thesis takes a risk – interpreters might not perform GA. If interpreters would remain visually engaged with notetaking, then the above overall research purpose would be forfeited. Therefore, some initial efforts are necessary to establish a case where interpreters would look away from their notepads during active listening and notetaking. The initial investigation is achieved via building up and coding a corpus. At Newcastle University, where this project is conducted, the Translation and Interpreting programme attracts over 80 students per annum. Each year, the trainee interpreters participate in CI examinations where the entire process is recorded. Although exploring eye movements usually involves the use of eye trackers, the CI final exam videos would prove valuable for the purpose of establishing a fact, as the abundant source material would contain rich data. Building up and coding the corpus generates abundant quantitative data, such as eye-movements patterns and textual features of the source speech. Analysing the data in a quantitative and inferential manner offers an opportunity for some general aspects of GA to be robustly explored (Bloomfield and Fisher, 2019).

The experimental study set out to explore GA after the corpus study has established the case that interpreters do perform GA during CI active listening and notetaking. Given the intricacy of the two topics and the scope of the study, it would be methodologically sound to conduct a follow-up study to verify the previous finding and explore the topic further and deeper. The experimental study involves manipulating variables to amplify the desired aspects for closer examination. Based on research aims and research questions, various types of data, such as GA frequency, GA length, and interpreters' performance score, are examined via quantitative and inferential analyses.

1.3. Structure of the Thesis

The remainder of this thesis is structured as follows:

Chapter 2 provides a review of relevant literature. The chapter consists of two parts. The first part reviews relevant literature on cognitive psychology with a special focus on cognitive load and GA. The second part of the chapter reviews relevant literature on CI active listening and notetaking.

Chapter 3 is an initial corpus study. 3.1 describes the specific aims and research questions of the corpus study as well as the overall methodological approach adopted in the study. Results obtained from the corpus study are reported in 3.2. Results are explained in 3.3. The chapter ends with 3.6, a summary of the study.

Chapter 4 is an experimental study that continues to explore GA in a CI setting. 4.1lays down the aims and research questions and describes the overall methodological approach used in the experimental study. 4.2 presents the results that are discussed in 4.3. The chapter ends with 4.4, a short summary of the study.

Chapter 5 is the General Discussion chapter, where overarching findings are discussed.

Chapter 6 concludes the thesis by summarising the main findings (6.1 – 6.4), rethinking GA as an indicator for high cognitive load and cognitive crisis (6.5), describing the strengths and limitations of the study (6.6), and, finally, providing directions for future research (6.7).

Chapter 2 Literature Review

Part I: Understanding Cognitive Load and Gaze Aversion

2.1 Human Cognition Architecture

2.1.1 Biologically Primary and Secondary Knowledge

Human beings are constantly bombarded with information from various sources, and processing information relies on a system of cognitive properties (Sweller, Ayres, and Kalyuga, 2011). According to Sweller *et al*. (2011), despite the numerous types of information that one could encounter, it can be coarsely divided into biologically primary and biologically secondary. Sweller *et al*. suggested a fundamental difference that distinguishes one from another: the manner of knowledge acquisition.

Biologically primary knowledge is not taught but acquired unconsciously and effortlessly for survival (Geary, 2007; Sweller *et al*., 2011). Such knowledge is the base from which a wide range of fundamental human skills are developed (Sweller *et al*., 2011). For example, most people pick up their first language without instruction, despite the fact that speaking involves a series of organs such as vocal cords, tongue, and lips. Sweller *et al*. (2011) proposed the modular nature of biologically primary knowledge, suggesting that skills-based on biologically primary knowledge are likely to have developed independently at different developmental phases. For instance, human beings can respond to face-like patterns soon after birth (Slater and Quinn, 2001; Kappas, Krumhuber, and Küster, 2013), whereas the skill of speaking comes at a later stage. Also, the cognitive mechanisms for the two tasks are expected to be different. Biologically primary knowledge, stored in long-term memory for easy and automatic application, is "a basic source for human ingenuity, creativity and skill", but it does not directly contribute to cognitive aspects that can be interpreted as intelligent behaviour because of how we express the biologically primary knowledge can be subject to "substantial alteration" (Sweller *et al*. 2011, p. 6). Using notetaking as an example, although the skill of doodling is acquired automatically and unconsciously, constructing a set of notes that suits the purpose of the task requires instructing and training. According to Sweller *et al*. (2011), if a skill must be altered to express its corresponding biologically primary knowledge, it would no longer be biologically primary but biologically secondary.

By contrast, obtaining biologically secondary knowledge requires effortful and conscious assimilation (Geary, 2007, 2008; Sweller *et al.*, 2011). Such knowledge is not learnt for the most basic survival purposes but for reasons such as advancing and perpetuating human civilisation. Scholars propose that compared with biologically primary knowledge, which is dealt with a specific cognitive medium, acquiring biologically secondary knowledge relies on a cognitive system, implying that acquiring biologically secondary knowledge carries a cognitive price.

2.1.2 Human Memory System

The notion that human beings have evolved the cognitive architecture where conscious and effortful processing of information is permitted towards a specific purpose is widely accepted among cognitive psychologists (e.g., Richard, Atkinson, and Shiffrin, 1971; Baddeley and Hitch, 1974; Sweller and Chandler, 1994; Baddeley, 2009; Moreno and Parl, 2010; Sweller, 2011). Over the past decades, cognitive psychologists have reached a consensus that the core of the cognitive architecture is the human memory system, a multi-modular construction that oversees not only storing but also processing and manipulating information (Atkinson and Shiffrin, 1968; Sweller, 1988; Baddeley and Hitch,1974; Baddeley, 2001; Cowan, 2005; 2010). So far, a wealth of literature has demonstrated the existence of a memory system consisting of Long-Term Memory (LTM), Short-Term Memory (STM), and Working Memory (WM) (e.g., Atkinson and Shiffrin, 1968; Baddeley and Hitch, 1974; Baddeley, 2009).

2.1.2.1 A Three-Store Memory System

Over the decades, a myriad of efforts has been invested in understanding how the human memory system is utilised when processing information. Atkinson and Shiffrin (1968) propose a model to illustrate the construction of the human memory system and its involvement in information processing. The model that has profoundly influenced the research agenda on memory consists of three stores: Sensory Register, Short-term Memory (STM), and Long-term Memory (LTM). The definitions and features of the three stores are summarised below:

2.1.2.1.1 The Sensory Register (SR)

The Sensory Register (SR) is a temporary store with an extremely limited capacity. The store scans and picks up the stimuli in the environment (such as sound and image) immediately once it is selected. The store can only hold information long enough to transfer it to the second primary component - STM, provided that the information is attended to.

2.1.2.1.2 Short-Term Memory (STM)

STM is a store where information can be retained temporarily. It is assumed that information, once entered the store as the result of attention, would decay and become lost entirely unless rehearsals are performed. The limited capacity of STM is believed to hold 5 to 9 digits ("*The Magical Number Seven*" (Miller, 1956) or last between 15 and 30 seconds. For example, someone may remember a phone number that he/she is about to dial, but the number will be forgotten within a few seconds. It is also suggested that the character of the information stored in STM may not necessarily mirror the form of the sensor input. For example, a visually exhibited word could be transferred from the visual sensory register into a short-term audio store.

STM, since introduced, has received extensive attention across disciplines, where immediate processing of information upon presentation is required (e.g., Carik and Lockhart, 1972; Brady, 1986; Tulving, 1985; Baddeley, 1986; 1992. 2012; Zen and Sak, 2015; Verhagen and Leseman, 2016; Pöchhacker, 2016; Gillies, 2017; Margo, Majerus, and Atout, 2020). The most influential theory developed from STM would arguably be Working Memory (WM) introduced by Baddeley and Hitch in 1974 (see 2.1.3 for detailed review).

2.1.2.1.3 Long-Term Memory

LTM is considered a major component within the model. Compared with the preceding two stores where information would decay and become lost, LTM holds information relatively permanently with an unlimited capacity. Information can enter LTM as a result of rehearsals which are controlled processes. Unlike the preceding two stores, whose capacities are limited to different degrees, LTM is understood to have an unlimited capacity (Atkinson and Shiffrin, 1968; Baddelely, 2009).

As reviewed above, STM has a very limited capacity and can only accommodate seven to nine digits or chunks. Chunking is the process of combining several items into one chunk, and it is LTM-dependent (Baddeley, 2009). Consider the following two sequences of letters: *GLKBQXNSGVMH* and *SPARTULUKOSA*. It is more likely that the latter would be easier to retain. The reason is that the language habits stored in LTM allow one to chunk the latter sequence into five chunks of syllables, whereas the first sequence is composed of unrelated letters and, therefore, difficult to chunk.

LTM consists of two distinctive storages, explicit or declarative memory and implicit or non-declarative memory (Graft and Schacter, 1985; Squire, 1992; Baddeley, 2009; Vakil, Wasserman, and Tibon, 2018; Wang, 2020). Explicit memory is where specific events and facts are stored, and it is explicit in a way that retrieving information needs conscious efforts, whereas implicit memory accommodates skills and procedures that one has learnt can apply subconsciously to perform a task, such as singing along a song or riding a bike (Cohen and Squire, 1980; Schacter, 1987; Baddeley, 2009). A recent study (Wang, 2020) shows that subjects would utilise both explicit and implicit memory stores in cross-situational word learning. Wang reported that explicit memory is involved in learning words related to familiar objects, possibly due to the available verbal encoding and retrieval strategy, whereas implicit memory supports learning words related to unverbalisable items, possibly resulting from absent verbal encoding and retrieval strategies. Conditioning, an associative learning procedure where a triggering stimulus is paired with a neutral stimulus to evoke a designated response (Baddeley, 2009), is found to be supported by implicit memory. For example, Weiskrantz and Warrington (1979) reported amnesic patients, who would suffer from semantic or episodic memory disorders, demonstrated conditioned response to the stimulus in anticipation. Implicit memory is also found to support priming, a process where the exposure to one stimulus influences the processing of a subsequent relevant item (Baddeley, 2009). Amnesic patients in word recognising tasks (Warrington and Weiskrantz, 1968) demonstrated that despite failing to retrieve information from semantic and episodic stores, they can take advantage of the priming. Race, Burke, and Verfaellie (2019) conducted a repetition priming experiment with amnesic patients and reported intact priming when associate learning is kept at a stimulus-response level.

Tulving (1972) suggests dividing explicit memory into semantic memory and episodic memory. Semantic memory is where someone's world knowledge, such as vocabulary and sensory tributes (taste and smell), are stored, and it is crucial for the use of language (Tulving, 1972). The storage of vocabulary is also called the mental lexicon (Vitevitch, Siew, and Castro, 2018). Apart from simple knowledge, semantic memory also houses schemata that are structures of domain-specific knowledge used to make sense of new material (Baddeley, 2009). Schemata play pivotal roles in affecting the amount of WM resources allocated to a task - the absence of appropriate schemata contributes to significant consumption of WM resources (Marcus, Cooper, Sweller, 1996). Baddeley (2009) suggests that schemata are beneficial as they help us form expectations for the world, enhance reading and speech comprehension, and perceive visual scenes.

Episodic memory receives and stores specific episodes or events and the spatiotemporal relations among them (Tulving, 1972, 1983; Baddeley, 2009; Sugar and Moser, 2019). In a recent study, Sugar and Moser (2019) postulate that episodic memory stores events in terms of location (where), sequence (when), and the contents (what). An interesting feature of episodic memory is that retrieving information from it, also known as "remembering or conscious recollection" or "mental time travel" (Tulving, 2002, p. 5), would create an input that keeps the episodes up to date, suggesting that episodic memory is receptive to change and information loss (Tulving, 1972).

Semantic memory and episodic memory are related. Tulving (1972) suggests that episodic memory operates based on, but goes beyond, semantic memory. Baddeley (2009) postulates that episodic memory can help form semantic memory in the sense that knowledge can be generated through experience. However, the two memory stores are vastly different (Tulving 1972, 2002; Wheeler, Stuss, and Tulving, 1997; Baddeley, 2009). Retrieving information from episodic knowledge, according to Wheeler *et al*., depends on a type of sensation that has been previously encountered. For instance, one can easily recall the taste of the spiced latté she or he had yesterday morning because the experience does exist. By contrast, retrieving information from semantic memory, such as recognising a low-frequency word, will not require someone traversing a specific sensational awareness. Semantic memory is also distinctive from episodic memory in a way that "it is a mental thesaurus, organised knowledge a person possesses about words and referents, about relations among them, and about rules, formulas, and the algorithm for the manipulation of these symbols, concepts, and relations" (Tulving, 1972, p. 386).

Based on the literature reviewed above, the construction of LTM can be illustrated in the following diagram:

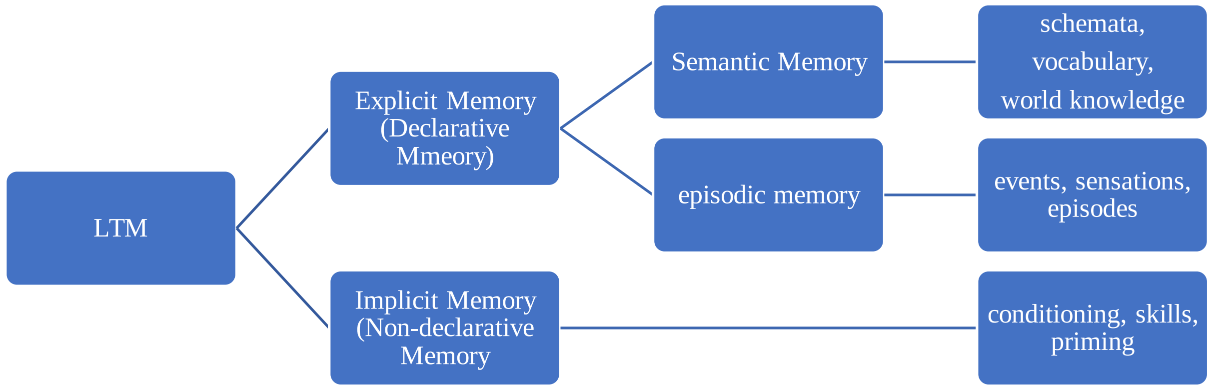


Figure 1 Construction of LTM

2.1.2.2 Controlled and Automatic Processing

Atkinson and Shiffrin (1968) also suggest that the three-store memory system works in a linear manner, a notion widely accepted by other cognitive scholars. The information flow is initiated with initial inputs into the Sensory Register. The next step is a subject-controlled scan within the sensory register, which could take place simultaneously with an associated search in LTM. The result of the preceding scan and memory search is the copy and transfer of the information from the Sensory Register to STM. Then, depending on the magnitude of the controlled process, such as rehearsing, the captured information could enter LTM. From the model, it is clear that although the process of newly selected information follows the linear order of the three stores, LTM can be involved throughout the process.

Richard *et al*. (1971, p. 82) proposed that the process in STM, that is, the flow of information into and out of STM during information processing, is under the "immediate control" of the individual. The level of such control will influence the extent to which information is stored in LTM. Common control processes or strategies include rehearsal (the repetition of information), coding (linking information with easily retrievable forms), and imaging (retaining verbal information through visual images). Although adopting the strategies is subject to multiple variables, they can be applied at the individual's discretion, which indicates that the controlled process requires cognitive efforts.

Schneider and Shiffrin (1977) made new contributions by proposing that automatic processing coexists with controlled processing and that the latter can be utilised to facilitate the former. Their findings were made possible through a series of studies measuring reaction time and accuracy in automatic detection and controlled search. According to Schneider and Shiffrin, human beings can learn how to process information automatically and store the skill in LTM. The involvement of LTM in information processing is later endorsed by other cognitive psychologists who subscribe to the notion that LTM accommodates domain-specific knowledge in the form of schemata (see 2.2.2.1) that help individuals to achieve efficient recognition and processing of the material (Baddeley, 1982; Sweller and Chandler, 1994; Schnotz and Kurschner, 2007). The automatic processing operates independently without taxing the attention capacity or using up STM resources. Automatic processing contains components that govern and manage attention, information flow, or response. Put simply, a person would conduct automatic processing effortlessly and subconsciously (Schneider and Shiffrin, 1977). Controlled processing, on the other hand, is a temporary activation of a sequence that has not been stored in LTM yet. Controlled processing taxes STM and requires attention from the person as it is initiated intentionally for a specific purpose. During controlled processing, one is aware of making efforts and is in control of evaluating the environment and choosing an appropriate strategy. Since controlled processing requires overt and effortful control, it would arguably be the case that activities where individuals are required to process presented information can be mentally challenging. For multitasking, in particular, mental demands can be extraordinarily high because the information can flow into STM from different sensory sources.

2.1.3 Working Memory (WM)

2.1.3.1 Overview

Working Memory (WM), as a system of cognitive properties, was brought forward by Alan Baddeley and Graham Hitch in 1974. However, the term was coined by Miller, Galanter, and Pribram in 1960 to refer to a mental faculty where information is kept for making plans. As mentioned earlier, Atkinson and Shiffrin (1968) also mentioned the term by equating it with STM, a unitary store that temporarily retains information. Baddeley and Hitch's WM is evolved from STM, suggesting similarities and differences between the two terms.

STM is a store capable of retaining information temporarily, whereas WM, according to Baddeley, is "a memory system that underpins our capacity to 'keep things in mind when performing complex tasks" (Baddeley, 2009, p. 9). In other words, STM deals with information retention while WM is involved in maintaining information while manipulating the information to support current cognitive activities.

Baddeley and Hitch (1974) offered two ways to interpret the application of WM, namely WMG, short for General Working Memory, and WMS, standing for Specific Working Memory. WMG refers to the storage that temporarily stores information that is being processed regardless of the range of cognitive tasks. By contrast, WMS depicts a detailed model of the structures and processes involved in performing cognitive tasks where WMG is required. In short, WM can be understood either as a processing capacity or a set of structures, each of which underpins cognitive tasks in a unique way. As a processing capacity, WM is believed to have a limited pool of resources, and substantial consumption of the capacity is likely to cause performances to deteriorate (Baddeley and Hitch, 1974; Gile, 1992; Baddeley, 2000, 2011; Cowan, 2009).

In terms of construction, WM is a two-hierarchy system with the Central Executive being the master component dominating a set of slave components which initially include the Phonological Loop and the Visual-spatial Sketchpad (Baddeley and Hitch, 1974). Baddeley (2000) introduced the third slave system, the Episodic Buffer. The Central Executive plays a dominant and supervisory role in the model. It controls cognitive processing by monitoring if the short-term storage is operating actively and intervening when the cognitive task is failing or when the person is distracted. The specific functions of the Central Executive include updating information, organising information from different sources, coordinating two assisting systems, operating selective attention, and facilitating shifts between tasks or strategies (Baddeley and Hitch, 1974; Baddeley, 2000; Wongupparaj, Kumari, and Morris, 2015). The Phonological Loop and Visual-spatial Sketchpad assist the Central Executive as two slave systems. While the Phonological Loop handles audio verbal information, the Visual-spatial Sketchpad stores and manipulates visual information (Baddeley and Hitch, 1974 Baddeley, 2000; Gluck, Mercado, and Myers, 2008). The Episodic Buffer not only serves as a buffer store that connects information across components of WM but also "links" WM to perception, LTM and semantics (Baddeley, 2000; Baddeley, 2011). It is also assumed that the Episodic Buffer could interact with taste and smell (Baddeley, Allen, and Hitch, 2011).

The proposal of WM laid the foundation for the emergence of other new theories (such as Cognitive Load Theory – see 2.2) and research in areas where multitasking is often required (such as Consecutive Interpreting - see 2.6). The following sections are dedicated to reviewing the literature on WM in terms of the nature of storage (unitary or multiple) and the construct. Studies on individual differences in WM, which have instantiated themselves during the corpus study (Chapter 3), will also be reviewed in this section.

2.1.3.2 Working Memory as A Limited capacity

There is a consensus among cognitive psychologists that WM has a limited capacity. But there is also an ongoing debate, that is whether WM is a system with a limited number of slots to accommodate items to be remembered (Miller, 1965; Chilchrist, Cowan, and Naveh-Benjamin, 2008; Cowan 2001, 2010) or is it a limited pool resource that can be simultaneously allocated to different concurrent tasks (Just and Carpenter, 1978, Alvarez and Cavanah, 2004; Wilken and Ma, 2004; Bays, Catalao, and Husain, 2009; Gorgoraptis, Catalao, Bays, and Husain, 2011). The former group of scholars believe that the increase of the number of items (quantity) that must be held in WM will cause performance to deteriorate, whereas their counterparts deem a sound allocation (quality) is the key to satisfactory performances. In spite of the disagreement, it is accepted by both sides that working memory has a limited capacity with distinctive individual differences (Baddeley, 2001; Cowan, 2005, 2010).

From the quantity perspective, like STM, WM is limited in the amount of information that can be held and in terms of duration of information retention. Miller, in 1956 published an influential article - "*The Magical Number Seven*" where he suggested that the immediate STM of adult human beings can only hold seven ± two pieces of meaningful information or chunks (Miller, 1956). Some later researchers cast different opinions on the exact number of pieces of information that can be retained in working memory. For instance, Chilchrist *et al.* (2008) demonstrated the amount to be three to four. In an effort to pin down the exact number of items that can be retained in working memory, Cowan (2010) introduced two distinctive measuring methods: process-related and storage-specific. According to Cowan, the former is used to test working memory ability, such as the extent to which an individual can adopt strategies to maximum performance (e.g., verbally rehearsing the items that need to be remembered). The latter, on the other hand, is utilised to test the capacity limits (e.g., a task where the items stored need to be evaluated while staying retained in memory). He concluded that for a young adult, the number of pieces of information that can be stored in central memory is limited to three to five (Cowan, 2010). The most obvious fact about the digit span measurement, a classic STM task where subjects are asked to recall number sequences with various digits immediately after the presentation, is that the limit for most people is about six or seven, but it could be shorter for words and even shorter if the words are foreign (Baddeley, 2009).

Another perspective on the limited capacity of Working Memory is from the resource allocation angle. Cognitive psychologists conceptualise WM as limited resource storage, and the resources will be simultaneously allocated to different tasks taking place concurrently. In a recent study, the researchers suggested that the allocation of WM resources is flexible and that it is the quality rather than the quantity of the representations (items stored) that determine the performance (Wei, Husain, and Bays, 2014). The notion about the limited capacity of WM, from the perspective of quality, is instantiated by the decay of accuracy in recall as the number of activities that consume working memory resources increase, hence the limited capacity of WM (Alvarez and Cavanah, 2004; Wilken and Ma, 2004; Bays, Catalao, and Husain, 2009; Gorgoraptis, Catalao, Bays, and Husain, 2011). The theories that look at WM as a limited capacity are stemmed from two premises: 1) Human beings' internal representation of sensory stimuli is "messy" because it can be featured with random and sudden fluctuations; 2) the increase in the number of stimuli can raise the mess level, and this rise is due to insufficient resources supply (Wei *et al*., 2014). Based on the two premises, it seems reasonable to suggest that if all the WM resources can be allocated to one task, ~~then~~ the performance of the task would be better than multitasking.

Psychologists who adopt the resource perspective also uphold the idea that the allocation of the resource is flexible. Bays and Husain (2008) advocated for a "dynamic shift" of limited working memory resources. They proposed that although the WM capacity is limited, it shifts among objects with flexibility. The allocation of resources is subject to selective attention and will be prepared for the items that come into visual attention. In other words, when we are moving our gaze to an item in the external stimuli, regardless of covertly or overtly, we are allocating resources to the item. This item is, therefore, retained in WM with higher precision (Bays and Husain, 2008). It has also been demonstrated that when channelling more resources to a prioritised item, the precision of recalling the item will be enhanced, but the recall of other items involved in the same task will be less successful (Gorgoraptis *et al*., 2011; Bays, Gorgoraptis, Wee, Marshall, and Husain, 2011), accentuating the nature of WM being a capacity with limited resources - if more resources are allocated to one task, there will be fewer resources available for other tasks.

2.1.3.3 Working Memory as a multi-component model

2.1.2.3.1 The Central Executive (CE)

The Central Executive is the dominating component of the WM system. Rather than being a memory system, it is an attentional controller that drives the system by allocating mental resources to the slave systems during cognitive tasks (Baddeley and Hitch, 1974; Baddeley, 1986, 2009, 2010, 2018).

What underpins the Central Executive is the two modes of controls proposed by Norman and Shallice in 1986. According to the authors, the first of the two ways of control is automatic control based on schemata stored in LTM and therefore consumes little attention. Another type of control, Supervisory Attentional System (SAS), on the contrary, requires mental efforts. SAS is believed to be a limited capacity triggered when automatic control falls short as a puissant attention regulator for tasks where splitting and mingling attention is inevitable (Norman and Shallice, 1986; Baddeley, 1986, 2009). In general, SAS will be summoned to play in a few scenarios: planning or decision making; problem-solving (especially when schemata are inadequate); processing novel or difficult information; handling danger or technical difficulties; relying on willpower (e.g., curbing temptation).

It is worth mentioning that the two control modes proposed by Norman and Shallice (1986) are much likened to what Schneider and Shiffrin (1977) submitted for automatic processing and controlled processing (see 2.1.2.1).

Owing to SAS's capability of the attentional control of action, the Central Executive is the most crucial component to WM and enjoys several capabilities. First, it facilitates focusing of attention, which is considered as the most essential function (Baddeley, 1986, 1996, 2002, 2009, 2010, 2018), allowing individuals to cut off from distractions. The second capacity of the Central Executive, according to Baddeley, is to divide attention between two or more tasks, allowing multitasking to take place. Given the focusing of attention function, the split attention shall be united when necessary. A third function that can be attributed to the Central

Executive is to shift attention from one task to another (Baddeley, 1986, 1996, 2002, 2009, 2010, 2018).

2.1.2.3.2 The Phonological Loop (PL)

According to the model, PL is one of the slave systems controlled by CE. It is assumed that PL is responsible for holding information in serial order and facilitating articulatory rehearsal (Baddeley and Hitch, 1974; Baddeley, 1986, 1996, 2002, 2009, 2010, 2018). PL has a limited capacity in storing and manipulating speech-like and acoustic information toward complex tasks. The storage can register about two digits or chunks of input for two seconds. Stored memory traces can be refreshed by subvocal rehearsal. Otherwise, they will decay and become lost. Another important function of PL is that it supports semantic coding, an encoding process in which the meaning of an item is processed and linked to LTM. PL is believed to play an influential role in linguistic activities such as speech comprehension and language acquisition. The reliance on PL is more significant when a subject is dealing with a foreign language than his or her native language.

Recently, Baddeley and Hitch (2019) published an update on PL as it is getting clear that the store does more than just temporarily retain verbal input. Studies show that articulation suppression in tasks of reading (Baddeley and Lewis, 1981) and distinguishing homophones (Baddeley and Lewis, 1981; Besner, Davies, and Daniels, 1981) would not lead to significant impairment to the performance. Therefore, it is suggested that PL might contain two underlying verbal codes, the articulatory auditory code that performs the original function of PL and a non-articulatory auditory code that underpins differentiating homophones but falls short of information retention or complex manipulation (Baddeley and Hitch, 2019).

2.1.2.3.3 The visual-spatial Sketchpad (VSSP)

Like the LP, the VSSP is another slave system under the CE. While LP is responsible for acoustic and audio input, VSSP is in charge of storing and processing visual and spatial forms of information (Baddeley and Hitch, 1974; Baddeley, 1986, 2009, 2010).

Logie (1995) proposed a two-component structure for VSSP, the visual cache and the inner scribe. The visual cache stores visual and spatial information passively, whereas the inner scribe refers to an active rehearsal process where visual-spatial inputs are manipulated towards complex tasks. It has been found that PL and VSSP can enhance cognitive performances when working together (Perason, Logie, and Gilhooly, 1999)

2.1.2.3.4 The Episodic Buffer (EB)

A fourth component, the Episodic Buffer (EB), was added to the model by Baddeley in 2000. Over the nearly twenty-five years after the original WM model, there had been evidence indicating interactions between WM and LTM. For example, PL can typically retain verbal information only for two seconds, but one can temporarily remember a short sentence with a duration exceeding the limit by using chucking, an LTM-based technique (Cowan, 2009; Baddeley, 2009). Besides, Baddeley (2000) also raised the question about the possible link between PL and VSSP - How is it that one can remember seven digits when PL can only accommodate two? If the rest is stored in VSSP, how are the two subsystems linked?

To provide answers to the questions, Baddeley (2000) postulated the existence of a buffer that allows multidimensional inputs (e.g., visual and verbal) from WM stores and LTM to gather and interact, linking WM to perception and LTM. This buffer is episodic because it binds information into several episodes or chunks. EB is found to be a limited capacity and can hold about four episodes (Cowan 2005; Baddeley, 2000). It is evident that binding information relies on CE to allocate attentional resources (Baddeley, 2000), but whether EB is a slave component under CE or a part of CE remains unclear (Allen and Baddeley, 2008).

2.1.3.4 Individual Difference in WM

Extensive efforts have been invested in addressing WM in terms of individual differences. Over the years, studies have found that people with higher WM span, compared with their counterparts, tend to do better in a variety of activities, such as reading comprehension (Daneman and Carpenter, 1980), syntax processing (King and Just, 1991), notetaking (Kiewra and Benton, 1988), writing (Benton, Kraft, Glover, and Plake, 1984), programming (Shute, 1991), following complex instructions (Engle, Carullo, and Collins, 1991), and fluid intelligence tasks (Engle, Tuholski, Laughlin, and Conway, 1999), among which the first three are more relevant to the general purpose of this project, i.e., exploring cognitive load and its indicator in Consecutive Interpreting (CI).

The trend of investigating individual differences in WM is sparked by Daneman and Carpenter (1980), who tested individual differences in WM with a reading span experiment that correlates with prose comprehension. A more specific aim of the study was to devise a way to operationalise the function of processing and storage of WM. In the test, participants were given a list of sentences to read. Afterwards, the participants were asked to recall the final word of each sentence in the order that the sentences were presented. The reading span of each participant was the "maximum number of sentences he or she could read while maintaining perfect recall of the final words" (Daneman and Carpenter, 1980, p. 452). Participants were also tested for their accuracy of comprehension of the sentences. The results indicate that performance in prose comprehension is positively correlated with WM capacity - compared with low span readers, high span readers were either more efficient in sentence processing or had a bigger storage capacity, or both (Daneman and Carpenter, 1980).

Another study that interests this project taps into individual differences in WM capacities correlating to syntax processing. Using the reading span test, Kings and Just (1991) recruited participants with high and low reading spans to participate in an experiment where syntactic structure and memory load were introduced as variables. Participants were asked to read sets of one, two, and three sentences and recall the last word of each sentence. Half of the sentences were subject-relative clauses, whereas the rest were object-relative clauses that scholars believe to be more cognitively challenging (Ferreira, Henderson, Anes, Weeks, and McFarlane, 1996; Gibson, 1998; Fedorenko, Givson, and Rhode, 2004; Fallon, Peelle, Wingfield, 2006). Memory load was manipulated by changing the number of the sentences preceding the final sentence based on which a true-false question was asked. The results showed that, for syntax processing, high span readers outperformed low span readers in accuracy and reaction time. Also, they reported that object sub-clauses were more difficult to process than subject sub-clauses, with the difference much more distinctive for low span readers. The findings provided evidence that working memory is taxed when processing syntax, and individual differences in WM capacity could be the key to the success or failure of processing complex syntax (King and Just, 1991).

Notetaking capabilities also prove to be subject to individual differences in WM capacity in terms of information processing ability (Kiewra and Benton, 1988). A group of undergraduate students were recruited to participate in the experiment. Before the experiment, the participants' information capability was assessed. They were given a lecture where they were asked to take notes. Kiewra and Benton hypothesised that the ability to process propositional information would be positively related to notetaking in terms of words, complex prepositions, and main ideas and that information processing ability can predict notetaking performance. The results indicate that the ability to hold and manipulate propositional information in WM is directly related to notetaking. Subjects who enjoy higher information processing ability would outperform their counterparts in terms of notetaking which is considered related to academic achievement. The author will continue to review the literature on syntax processing and notetaking in the realm of Interpreting Studies in 2.6.

2.2 Cognitive Load Theory (CLT)

2.2.1 Overview

Cognitive load refers to the demands on WM during cognitive tasks (Schnotz and Kurscher, 2007), and the theory evolving cognitive load (Cognitive Load Theory; CLT) is developed to understand human cognitive architecture and utilise the architecture to devise instructional procedures (Sweller, 2011). The inception of CLT began in the 1980sLarkin, McDermott, Simon, and Simon (1980) compared the strategy adopted by experts and novices when solving physics problems and reported that novices distinguish themselves from experts by adopting “means-ends analysis”, a conventional problem-solving strategy that requires the individual to work backwards from the goal to produce desired subgoals then reverse this procedure towards the goal state. Experts, on the contrary, worked directly towards the goal state. The advantage in efficiency, shown by experts, is the result of the utilisation of domain-specific knowledge in the form of schemata, a cognitive structure that allows individuals to recognise the problem state from previous knowledge and to pinpoint an appropriate strategy towards the goal (Sweller and Chandler, 1994; Schnotz and Kurschner, 2007). Sweller (1988) and his associates studied the “means-end analysis” strategy and found that although this conventional strategy helps to solve a problem, it may impose a heavy load on the cognitive capacity and therefore leaves little cognitive resources for schemata construction (Swellers, 1980, 1988, 1989; Moreno and Park, 2010).

CLT is developed based on the cognitive psychologists’ consensus that the human cognitive system contains a limited capacity known as working memory (WM) that is used in cognitive tasks (Miller, 1965; Baddeley and Hitch, 1974; Baddeley, 2000). WM is believed to be utilised to deal with novel information that is not yet stored in LTM in the form of schemata. However, the limitation of WM could make the assimilation of the new knowledge cognitively challenging (Sweller, 2005). The amount of WM resources dedicated to the assimilation of a particular material is understood as cognitive load. A high cognitive load, therefore, means that a large amount of WM resources is allocated to the learning task (Sweller and Chandler, 1994). For a learning task, the required volume of WM resources is largely influenced by whether corresponding schemata have been constructed and stored in LTM (Marcus *et al.*, 1996). According to Marcus *et al.*, successful learning requires the learner to comprehend all elements contained in the information. The required WM resource required by a particular learning task is, therefore, largely influenced by the volume of information that needs to be acquired and the extent to which the information can be recognised and processed in schemata. When appropriate schemata are present, the elements in the information can be processed as one unit, taxing a small amount of WM capacity and contributing to a low cognitive load. In contrast, in the absence of adequate schemata, the elements contained in the new information must be held and processed simultaneously in WM. Given the limited nature of WM, if the mental efforts required by the task surpasses WM capacity, a cognitive overload will be inevitable, and learning will be impaired.

Given that cognitive load describes the amount of WM resources expended on a task, it would be warranted to argue that tasks that rely on WM would induce cognitive loads to a degree. Typical activities that are WM-dependent include language comprehension, decision making, multitasking, and reasoning (Baddeley, 1992; Galy, Cariou, Melan, 2012).

2.2.2 Construct of Cognitive Load

2.2.2.1 Sweller's Model

Sweller and Chandler (1994) reported that cognitive load could be divided into intrinsic cognitive load and extraneous cognitive load. Intrinsic cognitive load is determined by the intrinsic complexity, that is, the level of elements interactivity of the material, and therefore cannot be altered. However, Brünken and his colleagues (2010) suggest that intrinsic cognitive load should not be defined by elements interactivity. They argued that the actual intrinsic load would be subject to the specific goal of the task, and therefore, the intrinsic cognitive load should be defined in terms of the amount of information that must be extracted from the information source (Brünken *et al.*, 2010). Extraneous cognitive load, on the other hand, is imposed by the manner of the presentation of the material, such as instructional design (Chandler and Sweller, 1991; Sweller, 1994). Since then, the CLT framework has emphasised the reduction of extraneous cognitive load in instructional design.

Studies on cognitive load in instructional designs are particularly keen on two learning effects, namely Split-attention Effect and Modality effect. Split-attention Effect refers to the situation where a learner divides his or her attention to mentally integrate various sources of visual information, such as understanding a geometry diagram while reading the statements. This mental combination is found to tax working memory (Tarmiri and Sweller, 1988; Chandler and Sweller, 1991, 1992; Yeung, 1999; Moreno and Mayor, 2000; Schnotz and Kurschner, 2007; Mayor and Logan, 2014; Schroeder and Cenkci, 2018). Modality Effect, in contrast, refers to the situation where a mixture of both verbal and visual information is present to the learner, such as using diagrams to explain abstract ideas. In the situation of a Modality Effect, the extraneous cognitive load is reduced and, therefore, is preferred (Watkins and Watkins, 1988; Mayor, 1997, 2001; Tabbers, Martens, and Merienboer, 2001).

Although CLT was developed to achieve sound instructional designs to promote learning, it was not clear how or why learning is attained until the late 1990s when Sweller and his colleagues (1988) introduced a third type of cognitive load, that is, germane cognitive load. Germane cognitive load is imposed by devoting efforts towards schemata acquisition or automation and therefore can facilitate learning to take place. The devoting efforts typically include conscious activities that allow learners to apply learning strategies, link the material with existing schemata, restructure representation of the problem, and trigger meta-cognitive processes (Schnotz and Kurscher, 2007). Therefore, while intrinsic cognitive load and extraneous cognitive load are imposed as a result of the task requirements, germane cognitive load, however, is imposed by activities that go beyond the task requirements and aim at learning. Put simply, intrinsic and extraneous cognitive load are concerned with task performance, whereas germane cognitive load is learning-based (Sweller *et al.*, 1988; Schnotz and Kurscher, 2007; Galy *et al.*, 2012).

The three types of cognitive load share the same WM capacity, and the combined level of the three types of cognitive load are additive and must not exceed the unused WM capacity to establish schemata acquisition or automation (Paas, Renkl, and Sweller, 2003, 2005; Ginns, 2006). For purposes other than learning, such as solving mathematical problems or regulating air traffics, good performances would require the combined intrinsic and extraneous cognitive load to remain within the capacity of WM (Sweller *et al.*, 1988; Schnotz and Kurschner, 2007; Galy *et al.*, 2012).

2.2.2.2 Paas and Merriënboer's Model

Although the three types of cognitive load are widely accepted among scholars, the interpretations of the construct of cognitive load differ. Paas and Van Merriënboer (1994) proposed a model to visualise the construct of cognitive load. According to the model, cognitive load has two dimensions, namely, assessment factors and causal factors.

The assessment dimensions of cognitive load include three factors, namely, mental load, mental effort, and performance. The mental load is imposed from the complexity of the task. Mental effort refers to the amount of WM resources allocated to accommodate the task demands. The performance element can be used to deduce the level of cognitive load - smooth performance with less mental effort would indicate a low level of cognitive load.

The causal factors indicate the interactions between the task characteristics and the individual's (such as the learner's) characteristics. According to the authors, the task characteristics describe both the features of the task and the learning environment. However, Choi, Merriënboer, and Paas (2014) proposed an updated model where the characteristics of the physical environment are separated from the task features, suggesting that cognitive load can be induced by the factors in the environment. Previous studies have discovered that distracting visual stimuli in the environment can hinder task performance, and subjects would avert gaze (see 2.4.4) to disengage from distractions (Glenburg, Schroeder, and Robertson, 1998; Doherty-Sneddon, Bruce, Bonner, Longbotham, and Doyle, 2002; Doherty-Sneddon and Phelps, 2005; Doherty-Sneddon, Riby, and Whittle, 2012). Choi *et al*. proposed that extraneous cognitive load can be induced by environmental influences that should be minimised when possible so that more resources can be devoted to the task.

2.2.2.3 Cognitive Load in multitasking

Although CLT was initially developed to guide instructional designs, it has been addressed in other areas outside the realm of learning.

Complex tasks that require one to perform several tasks concurrently impose a high cognitive load. For example, some ergonomists studied working environments where multitasking should be performed and found that various factors would result in cognitive overload, which would impede performance or even safety (Hart and Staveland, 1988; Reid and Nygren, 1988;

Miyake, 2001; Galy, Cariou, Melan, 2012). The gross cognitive load imposed by the task and the environmental demands is considered a mental load (Kablan and Erden, 2007; Kirschner, 2002). Galy *et al.* (2012) tested the additive interaction between intrinsic, extraneous and germane cognitive load by manipulating mental workload factors in a WM task and revealed the existence of an additive effect between task difficulty and time pressure. with the former inducing intrinsic cognitive load and the latter imposing extraneous cognitive load. According to the authors, high time pressure would induce anxiety which costs more WM resources to be channelled to the task. Bong, Fraser, and Oriot (2016) reported that stress contributes to extraneous cognitive load and, at a certain threshold, can impair WM performances. Galy *et al*. proposed that the remaining WM resources, that is, the available WM capacity that has not yet been consumed by task difficulty and time pressure, would allow individuals to apply strategies to achieve good performance and therefore is germane.

Several studies have found that work environments where noise or irrelevant auditory and visual stimuli are likely to burden WM (Murphy, Craik, Li, and Scheider, 2000; Speranza, Daneman, and Schneider, 2000; Johnson and Zatorre, 2005; Heinrich, Schneider and Craik, 2008; Anderson and Kraus, 2010; Coffey, Mogilever and Zatorre, 2017). Noise has been found to impair STM tasks by taxing processing resources or hindering sensory representation, implying that unnecessary extraneous cognitive load can be imposed by noise (Murphy *et al*.,2000; Heinrich *et al*., 2008). For listening tasks, splitting useful information from background noise requires the perceiver to store the selected information in WM while ignoring distracting noise by allocating attentional resources that otherwise would have been devoted to speech processing (Anderson and Kraus, 2010). Visual input can also be distracting. In an fMRI study, Johnson and Zatorre (2005) found that when participants were asked to simultaneously process unrelated auditory and visual input, although BOLD response increased in sensory cortices corresponding to the stimulus that participants focused on and decreased in the sensory cortices corresponding to the ignored stimulus, a certain volume of attentional resources is consumed by the irrelevant input. The human face encompasses rich signals and can add difficulty to cognitive tasks (see 2.3.2.2).

Xie and Salvendy (2000) proposed a detailed model, another perspective to understand the construct and measurement of cognitive load in a multitasking working environment. The model contains five types of workloads. *Instantaneous Workload* refers to the dynamics of workload that fluctuates during a task and serves as a basis for other workload types. *Peak Workload* is defined as "the maximal value of instantaneous mental workload when performing a task" (p. 88). According to the authors, the instantaneous workload must remain below one's mental-load limit to avoid cognitive overload and performance deterioration. The third type of workload in the model is *Accumulated Workload*, meaning the combined workload experienced, or the total amount of information processed during a task. The fourth attribute is *Average Workload*, defined as "the average value of instantaneous workload" and is considered to "equal the accumulated workload per unit time" (p. 89). It is the Average Workload that indicates the mental intensity of a task. Finally, the fifth type is *Overall Mental Workload*, which is used frequently in subjectively measuring cognitive load. Overall mental load indicates the individual's experience of mental workload based on the whole task procedure or the combination of instantaneous, accumulated, and average workload.

Based on the existing studies, the following is drawn to compare the perceptions of three types of cognitive loads in the scenarios of learning and multitasking.

Table 1 Three types of cognitive load in learning and multitasking

	Learning	Multitasking
Intrinsic cognitive load	• Imposed by the inherit difficulty from the learning material (such as a geometry problem); • Cannot be altered; • Additive with extraneous cognitive loads	• Imposed by the inherit difficulty of the task (such as air control); • Cannot be altered; • Additive with extraneous cognitive loads
Extraneous cognitive load	• Imposed by the manner of presentation of the material and activities that required to perform the task; • Can be regulated by improving the instructional design; • Additive with intrinsic cognitive load;	• Imposed by time pressure, anxiety, stress, audio and visual distractions; • Can be altered by allowing a wider time frame or cutting off from distractions; • Additive with intrinsic cognitive load;
Germane cognitive load	• Subject to the availability of free WM capacity;	• Subject to the availability of free WM capacity;

	• Imposed by conscious activities aimed at learning (such as applying appropriate strategies, reconstructing the representation of the problem, abstracting schema from LTM, and meta-cognitive monitoring) (Schnotz and Kurscher, 2007, p. 497);	• Imposed by the conscious application of strategies to perform the task;

It should be noted that despite the varying interpretations of how cognitive load is constructed, scholars concerned with CLT seem to have reached a consensus that the level of cognitive load is determined by the amount of WM resources invested in the task process.

2.2.3 Measuring Cognitive Load

Despite the advancements in CLT in the past decades, there has not yet been a standardised or common cognitive load measurement method (Brünken, Seufert, and Paas, 2010). Paas and van Merriënboer's model suggests that there are three measurable dimensions of cognitive load, namely mental load, mental effort, and performance (Paas and van Merriënboer, 1994a; Kirschner, 2002; Kablan and Erden, 2007; Galy *et al*., 2012). Mental load, a task-centred dimension, refers to the cognitive load solely imposed by the task (Choi, Merriënboer, and Paas, 2014). Mental effort, a human-centred dimension, describes the amount of WM resources allocated to deal with the task (Choi, *et al*., 2014). The performance of the individual can help infer the level of cognitive load (Kirschner, 2002; Choi, *et al*., 2014).

The cognitive load measurements methods can be divided into three categories – subjective, objective, physiological, behavioural, and efficiency measurement.

2.2.3.1 Subjective Measures

Subjective methods are based on the assumption that individuals can introspect on their cognitive process and estimate the extent to which WM was taxed in a task. Such methods, which utilise self-reported rating scales to assess mental effort, are the most common ways of measuring cognitive loads (Brünken *et al*., 2010). Subjective measurements usually include subjective ratings of perceived mental effort and task difficulty.

Subjective measures typically require participants to rate their perceived cognitive load, which is often combined with the rating of perceived task difficulty, by using the scale ranging from 1 ("very, very low") to 9 ("very, very high") (Paas and van Merriënboer, 1993, 1994b; Paas, Tuovinen, Tabbers, and Van Gerven, 2003). Depending on the aim of studies, researchers can either adopt unidimensional scales to investigate a single variable or use multidimensional scales that assess a group of associated variables such as mental effort, fatigue, and anxiety (Paas *et al*., 2003; Chen, Zhou, Wang, Yu, Arshad, Khawaji, and Conway, 2016). Brünken *et al.* (2010) summarised that despite the apparent benefit of subjective methods, that is, simplicity, there are several limitations. First, subjective ratings usually offer a post hoc general assessment cognitive load induced by the learning or working task. Therefore, unless conducted repeatedly at different stages, the results can not specify whether the process or the task's extra activities have caused the cognitive load to increase. Second, subjective measures neither indicate which of the three types of cognitive loads are assessed nor reveal which type of the cognitive loads originated the reported mental efforts.

2.2.3.2 Objective Measures

Apart from subjective measures, cognitive loads can also be assessed with objective methods that cover the measuring of the outcome, task complexity, and behavioural indicators, including neurophysiological measures, time-on-task, information retrieval, and dual-task (Brünken *et al*., 2010). The measuring of the outcome or performance is based on the basic CLT notion that the outcome of a task (e.g., learning) is negatively correlated with the cognitive load level (Sweller and Chandler, 1994; Sweller, 2011; Schroeder and Cenkci, 2018). Therefore, an unsatisfactory outcome of performance could indicate a high cognitive load or overload. However, the cognitive load assessed by merely measuring the outcome can be unreliable because other variables, such as motivation, could affect the outcome (Brünken *et al*., 2010).

Similarly, assessing cognitive load via measuring task complexity can be grounded in CLT - intrinsic cognitive loads are the results of the task's inherent complexity of the task (Sweller *et al*., 1994). Therefore, the task complexity level would indicate the necessary WM resources that are devoted to the task. However, this method can be unreliable. Although intrinsic cognitive loads are imposed by the task's inherent complexity, whether or not the participants experience the maximum cognitive load that the task can impose will depend on the purpose of the task (Brünken *et al*., 2010). If the task is to memorise the first and the last word of an abstract theory speech, the cognitive load for this particular goal would be lower than if the goal is to assimilate the knowledge.

2.2.3.3 Physiological Measurements

Compared with task performance and complexity, physiological measures assume that functional physiological variables would indicate the fluctuation of cognitive load. Such measures can show the real-time changes of cognitive loads as behaviours are more directly related to the task (Paas *et al*., 2003; Brünken *et al*., 2010). One of the physiological techniques is Functional Magnetic Resonance Imaging (fMRI) that monitors the Blood Oxygenation Level Dependent (BOLD) in different areas in the brain during a cognitive task (Whelan, 2007). The rationale behind this method is that active areas of the brain during cognitive tasks attract more oxygenated blood, and such oxygen consumption can be detected in magnetic fields generated by an MRI scanner. For example, during cognitive tasks, the brain areas that are more active, i.e., correspond to higher neural activity, would attract higher levels of oxygenated blood. Using the colour coding method, the fMRI technique can demonstrate the increase or decrease of oxygenated blood, from which the change of cognitive loads can be deduced (Whelan, 2007). Although the fMRI technique could demonstrate continuous fluctuation of cognitive load, the application of such technology requires a sophisticated, costly, and bulky apparatus and therefore is used in highly specialised laboratory environments (Whelan, 2007; Brünken *et al*., 2010). Other available physiological parameters typically involve measuring heart rates and tracking eye movements which are also based on the assumption that the changes in cognitive functioning could be mirrored in the changes in physiological functioning. Studies have found that heart rate variability (HRV) is positively correlated with task demands (Mulder and Mulder, 1981; Paas and van Merriënboer, 1994b; Tattersall and hockey, 1995; Haapalainen, Kim, Forlizzi, and Dey, 2010; Galy *et al*., 2012; Cranford, Tittmeyer, Chuprinko, Jordan, and Grove, 2014). However, Paas and van Merriënboer (1994b) reported that HRV is insensitive to subtle cognitive load fluctuations. By contrast, the pupillary response has been found to be a highly sensitive indicator for the fluctuation of cognitive load. Beatty and Lucero-Wagoner (2000) identified three task-evoked pupillary responses (TEPRs), i.e., mean pupil dilation, peak dilation, and latency to the peak. Van Gerven, Paas, van Merriënboer, and Schmidt (2004) measured cognitive pupillary response in a Sternberg memory-search task and found that mean pupil dilations of young adults are considerably sensitive to memory loads.

2.2.3.4 Dual-Task Measurement

Another type of measurement method is dual-task measurement. As previously mentioned, primary task measurements help to indicate cognitive load imposed during a task by assessing how well the task was performed. Secondary task measurements, also known as dual-task measures, take advantage of the limited capacity of WM by asking the participant to concurrently perform two tasks with the second task being simpler than the first one but still requires sustained attention (Paas and Van Gerven, 2003; Brünken *et al*., 2010). The idea behind this method is that when the first task (such as understanding a presentation) becomes cognitively demanding, it would be challenging to maintain the second task (such as pressing the space bar on the keyboard when a signal is given while performing the first task) (Moreno and Mayer, 2000; Brünken, Steinbacher, Plass, and Leutner, 2002; Brünken, Plass, and Leutner, 2004). There has not been a broad application of the dual-task measurement method due to its serious disadvantages. To begin with, the second task could considerably interfere with the first task, especially when the first task imposes a high level of cognitive load (Paas and Van Gerven, 2003). Also, although the general level of cognitive load can be deduced, this comparative method cannot provide an accurate estimation of resource consumption (Brünken *et al*., 2010).

2.2.3.5 The Efficiency Measurement

So far, this section has reviewed subjective, objective, physiological, and behavioural, cognitive load measurements that could indicate imposed cognitive load by assessing an independent variable. As reviewed above, despite the efficacy in indicating cognitive load, each of the measurements has shown some degree of disadvantages. A shared shortcoming among these measurements is that individual measures targeted at one variable (such as performance) cannot guarantee comprehensive and meaningful interpretations of the interaction between a specific task and induced cognitive loads because the variable can be confounded with other variables (Paas and Van Gerven, 2003; Brünken *et al*., 2010). For instance, for tasks like air traffic control and interpreting, circumventing a high cognitive load is not as crucial as achieving good performance, which is usually expected. In these situations, good performance does not equal low cognitive costs but indicates high alignment with the task goal. Attempting to address this issue, Paas and van Merriënboer (1993) proposed an approach to demonstrate the relation between mental efforts and performance by yielding an efficiency score via calculation using the following formula:

$$\boldsymbol{E} = \frac{z_{Performance} - z_{Mental\ Effort}}{\sqrt{2}}$$

$\boldsymbol{E}$ stands for efficiency, $z_{Performance}$ and $z_{Mental\ Effort}$ stand for the score for performance and mental effort, respectively.

The efficiency measurement takes both cognitive load and performance into consideration and can be applied to situations where one element outweighs another or several variables are present (Brünken *et al*., 2010). Instead of emphasising the maintenance of a level of cognitive load that is manageable to participants, the efficiency measurement advocate for the situation where high-task performance is associated with low cognitive effort (Paas and van Merriënboer, 1993).

2.2.3.6 Other Measurements

As mentioned earlier, cognitive loads are observed and addressed in multitasking environments. Cognitive loads measurement methods in work environments can be divided into subjective measures, performance measures, and psychophysiological measures (Galy *et al*., 2011). Subjective measurements include the NASA Task Load Index (NASA-TLX) and Subject Workload Assessment Technique (SWAT). The two methods evaluate factors such as task difficulty and time pressure that could contribute to cognitive overload and hinder performance. Performance, response accuracy and response latency, in particular, can also indicate cognitive load. However, the performance measurements would rely on a number of behaviours, and they could fall short of indicating subtle changes in cognitive loads (Galy *et al*., 2011). Similar to the situation of learning, the efficiency measurement (Paas and van Merriënboer, 1993), therefore, is preferred when both cognitive load and performance are taken into consideration.

The fluctuation of cognitive loads can also be indicated by eye movements. Studies have found that during conversations, one would typically look away from the interlocutor when the question is difficult because the human face can signal rich information and, therefore, can be distractive (Doherty-Sneddon *et al.*, 2002). Earlier studies suggested a cognitive load hypothesis that averting gaze is faciliatory as it helps individuals to shut down from the distractions in the environment and allows more mental resources to be channelled to the task, and therefore enhances performance (Glenburg, Schroeder, Robertson, 1998). To understand the function of looking away during cognitive tasks and potentially explore it in the sphere of interpreting studies, it would be warranted to gain a full picture of the human gaze.

2.3 Human Gaze

2.3.1 Overview

A wealth of studies has accentuated the facilitatory role of visual communication signals such as body language and facial expressions in human interactions (Clark & Brennan, 1991; Goldin-Medow, Wein, and Chang, 1992, McNeil, 1985; Doherty-Sneddon *et al.*, 2002). Since the 1960s, much attention has been invested in understanding the human eyes. As the media for both signalling and perceiving information, eyes are believed to bear irreplaceable significance in social behaviours (Argyle and cook, 1976; Argyle, 1988; Gobel, Kim, and Richardson, 2015).

Evidence suggests that human beings start to engage in gazing behaviours at the early stage of life (Farroni, Csibra, Simion, and Johson, 2002). By the age of two months, human infants would prefer to gaze at eyes rather than other regions of human faces (Hainline, 1978; Haith, Bergman & Moore, 1977; Maurer, 1985). Three to six months old infants would be able to follow an adult's ostensively direct gaze to an object (Hood, Willen, and Driver, 1998; Senju, Csibra, 2008). By the age of four months old, infants can tell the difference between direct and averted gaze (Johnson and Vecera, 1993; Vecera and Johnson, 1995). Five-month-old infants would look longer at a face from which a direct gaze has been previously detected (Lasky and Klein, 1979). Nine-month-old infants can establish shared foci of attention with their carers (Scaife and Bruner, 1975). For children, being attentive to gaze forms a base for empathy and conscience development (Dadds, Allen, Oliver, Faulkner, Legge, Woolgar, and Scott, 2012).

The amount of gazing behaviours is found to increase in childhood (four to five years old) and decline during adolescence and build up again in adulthood (Levine and Sutton-Smith, 1973). Scholars on human gaze patterns suggested that the behaviour is subject to considerable individual differences in both social settings (e.g., Rogers, Speelman, Guidetti, and Longmuir,

2018) and situations where monitoring face (e.g., viewing images) on a screen is required (Coutrot, Binetti, Harrison, Mareschal, and Johnston, 2016; Arizpe, Walsh, Yovel, and Baker, 2017). Based on the aove According to Rogers *et al*., memory skills or the function or purpose of a specific setting could contribute to individual differences in gazing pattern. Gaze behaviours also vary among cultures. While some cultures favour eye contact during social interactions (e.g., conversation), some would encourage otherwise (Argyle, 1988; Bohannon, Herbert, Pelt, and Rantanen, 2013). For example, Argyle reports that Arabs, Latin Americans, and Southern Europeans make more eye contact than British, White Americans, or Asians (Agyle, 1976). Japanese culture, as another example, is a typical non-contact culture where eye contact is discouraged as it is considered disrespectful, and therefore people tend to look at others' necks so that the visual attention will still fall in others' peripheral vision (West, 1995; Morsbach, 1973; McCarthy, Lee, Itakura, and Muir, 2006, 2008). Moreover, in Japan, if social rank matters during social interactions, one would be advised to avoid eye contact with the superior (West, 1995).

2.3.1.1 Gaze in Social Interactions

As mentioned earlier, eyes are dual-functional (Gobel *et al*., 2015). As an outlet channel, eyes can convey rich emotions and interpersonal attitudes, such as excitement, sadness, fear, anger, and shame (Tomkins and McCarter, 1964; Adolphs, Gosselin, Buchanan, Tranel, Schyns, and Damasio, 2005). However, one of the most significant functions of the gaze is to indicate liking (Argyle and Cook, 1976; Klienke, 1986; Aygyle, 1988). For example, Exline and Winters (1966) investigated the "look-like" relationship with an experiment where subjects were either insulted or praised by the same confederate and reported that the insulted subjects looked much less at the confederate than those who received compliments. The result is endorsed by other studies that report similar results that people tend to pay more visual attention to those whom they like and gaze less at whom they disapprove (e.g., Efran, 1968, 1970; Mehrabian, 1972, 2017; Straaten, Holland, Finkenauer, Hollenstein and Engels, 2010; Adams, Nelson, and Purring, 2013). The function of gaze in signally liking is elaborated further by Argyle and Dean (1965), who include eye contact as the first and foremost ingredient in constructing intimacy. Gaze can signal the opposite of liking, such as hostility and aggression (Ellsworth, 1975; Exline, Ellyson, and Long, 1975). However, aggression and hostility can be delivered either by staring (such as hate stare) or looking away (i.e., ignoring) (Argyle and Cook, 1976).

Apart from signalling attitudes and emotions, the gaze also serves as a channel through which human beings can ascertain others' mental and emotional states (Baron-Cohen, Wheelwright, and Jolliffe, 1997; Adams *et al*., 2013; Gobel *et al*., 2015). The facial structures surrounding the eyes are believed to be able to facilitate nonverbal communications (Ekman and Friesen, 1975). Bruce and Young (2012) suggested that the human face can indicate various signals that are usually correctly interpreted by others. Human beings tend to pay attention to others' faces readily and are more skilled at detecting faces than other objects (Hershler and Hochstein, 2005). In particular, human beings tend to fixate on the eyes during face-to-face interactions (Bruce and Young, 2012; Adams *et al*., 2013). Direct gaze signals intention towards the perceiver, conveys essential social and communication information and therefore is predicted to capture the perceiver's attention and make it difficult to disengage, leading to mutual gaze (Senju and Hasegawa, 2005).

Mutual gaze is important in face-to-face communications, and it would become more frequent as the distance between interlocutors increases (Argyle and Dean, 1965). Kleinke (1986) reported that slightly longer than usual mutual gaze indicates the listener's turn to speak, whereas quick mutual glances could mean approval and understanding. Even in virtual settings, such as video conferences, attendees would look into the eyes on the screen rather than the camera sitting on top of the monitor (Bohannon *et al*., 2013). Mutual gaze can help detect friendship and establish trust (Nurmsoo, Einav, and Hood, 2012; Bekkering, 2004; Bohannon *et al*., 2013).

A significant role of mutual gaze lies in its cueing effect, where another person's gazing can be puissant in orienting other's selective attention (Friesen and Kingstone, 1998; Driver, Davis, Ricciardelli, Kidd, Maxwell, and Baron-Cohen, 1999; Langton and Bruce 1999; Kuhn, Tatler, and Cole, 2009). Kuhn *et al*. (2009) conducted an interesting experiment where a magic trick is presented. During the magic show, participants' attention was manipulated by the magician's gaze direction. Once the magician had visually endorsed the misdirection, participants became significantly less likely to detect the to-be-concealed event. The experiment also showed that participants gazed less at the critical hand when the magician used gaze to misdirect their attention to the other hand. The last finding of the study is that the magician's face, in particular, the eyes, attracted most of the fixations from the participants. This cueing effect can also lead to the subjective evaluation of objects, such as liking (Grynszpan, Martin, and Fossati, 2017). In a study, participants were found not only to follow others' gaze but also to show liking towards the objects that had previously received visual attention (Bayliss, Paul, Cannon, and Tipper, 2006).

Given the significance of gaze, failure to engage in gaze may hinder communication (Fullwood and Doherty-Sneddon, 2004) and is considered a sign of social incompetence (Argyle, 1988). Failure to develop typical mutual gaze behaviours is one of the earliest signals for severe social and communicative disorders (such as autism) or even sociopathic personalities (Volkmar and Mayes, 1990; Baranek, 1999; Pellicano and Marcae, 2009; Dadds *et al.*, 2012).

2.3.1.2 Gaze in Cognitive Activities

Apart from social interactions, human gazing behaviours have been found to reveal aspects of cognitive activities, too. For instance, in classroom settings, it is reported that if the teacher gazes at the students frequently, students will remember more instructions (Fry and Smith, 1997). This memory-enhancing effect is because that gaze can serve as an arousal stimulus that attracts attention, thereby boosting memory (Kelly and Gorham, 1988).

In the past decades, cognitive psychologists have conducted studies featuring eye-tracking methods to observe eye movements correlating to cognitive efforts in various cognitive activities. The eye-tracking approach is well-grounded in the gaze-comprehension model (Just and Carpenter, 1980), which hypothesised that the mind comprehends visual inputs.
The literature has shown a fashion in exploring two eye movements: fixations and saccades. Fixations refer to "eye-movements that stabilise the retina over a stationary object on interest" (Duchowski, 2017, p. 44), whereas saccades are defined as the "eye movements used in repositioning the fovea (the centre of the retina with the highest visual acuity) to a new location in the visual environment" (Duchowski, 2017, p. 41).

Eye fixations and saccades have received much attention in understanding reading (e.g., Just and Carpenter, 1980; Liversedge and Findlay, 2000). Eye fixations during reading are word-based (Just and Carpenter, 1976; Radach and Kennedy, 2004). In most cases, saccades in reading would depart from one word and land on the same word or the onwards (Radach and Kennedy, 2004), although readers sometimes regress (Jarodzka and Brand-Gruwel, 2017). Studies have found that readers tend to fixate less on or skip words that are short, common, easy to predict, and lexically unambiguous (Altaribba, Kroll, Sholl, and Rayner, 1996;

Brysbaert and Vitu, 1998; Drieghe, Brysbaert, Desmet, and De Baecke, 2004). By contrast, uncommon words would associate with longer fixations due to the more significant processing load (Just and Carpenter, 1980; See 2.6.2.4 for more detailed review regarding uncommon words). Fixations also tend to be shorter when the reader is less mindful, as mindless reading requires less cognitive effort and decouples eyes from text processing (Reichle, Reineberg, and W. Schooler, 2010).

Although eye movements have been frequently measured to study reading, attention is increasingly given to spoken language comprehension. Like the notion that fixations are word-based in reading, eye movements in spoken language processing are closely time-locked to the input (Tanenhaus, Magnuson, Dahan, and Chambers, 2000). Tanenhaus *et al.* (2000) tracked participants' eye movements as they listened to a group of sentences, one of which contained syntactic ambiguity while looking at displays of pictures that the sentence referred to. The sentences used in the experiment were A) "Put the apple on the towel in the box", and B) "Put the apple that's on the towel in the box". Sentence A is ambiguous as the prepositional phrase "on the towel" would initially be processed as the destination for the apple to be put, but as soon as "in the box" was uttered, a garden-path effect was observed. By contrast, sentence B made it clear that the "box" is the destination of the apple. Data from the eye-tracking showed that fixations on the false goal were significantly more often when processing the ambiguous sentence than the unambiguous sentence.

However, people are more attuned to process abstract spoken language and references without using visual stimuli. In a study, Heutte, Winter, Matock, and Ardell (2014) postulated that eye movements, saccades and fixations would mirror the underlying conceptual structure of linguistic features in the absence of visual stimuli. Specifically, mentally processing a verbal sentence describing motion (e.g., “*he was walking*”) would result in more dynamic eye-movements in contrast to mentally processing a simple past sentence (e.g., “*he walked*”) would lead to less eye-movements. According to Huette *et al.* (2014), this result would indicate that grammatical differences in linguistic input can induce eye-movements.

With several concurrent subtasks vying for limited attentional resources, interpreting is a hardcore activity believed to be fundamentally challenging (Danks, 1997; Shreve, 1997; Nigro, 2015). O'Brien (2007) pioneered using eye-trackers in translation and studies when she explored the levels of the cognitive load imposed by different Translation Memories in machine translation. For interpreting, it was not until recently that scholars began to use eye-tracking to investigate how interpreters' eye movements might indicate the fluctuation of cognitive load (e.g., Seeber, 2015; Seeber and Kerzel, 2012; Stachowiak-Szymczak, 2019; Lin, 2022). Eye-tracking seems to fit in the study of Sight Translation as the task requires an interpreter to orally render a written document in a target language with a short eye-voice span (Mellinger, 2017). Dragsted and Hansen (2009) suggested that interpreters usually fixate on the texts under processing during Sight Translation, whereas translators working on the same material tend to fixate on different locations. According to Dragsted and Hansen, this difference means that fixation patterns can indicate the distribution of attentional resources during text-based translation tasks. Chmiel and Mazur (2013) reported that mental efforts in Sight Translation are indicated by fixation duration that negatively correlates with sentence readability. Su (2020) conducted an eye-tracking study focusing on interpreters' fixations and saccades during preparations and executions of Sight Translation. She reported that preparations were associated with fewer and shorter fixations and saccades, whereas the actual translation leads to significantly longer fixations and noticeable pupil dilations, which, according to Su, is due to the high cognitive load resulting from having to multitask. Stachowiak-Szymczak (2019) adopted the eye-tracking method in studying both simultaneous interpreting (SI) and CI and suggested that eye movements correlate with language processing during interpreting. In the CI part of the study, the source texts were controlled with numbers and lists as problem triggers. Interpreters were allowed to take notes but to look at pictures that were either congruent or incongruent with the source speech or a black screen so that a high memory load was maintained during listening. Stachowiak-Szymczak reported fixation durations were the longest when interpreters process problem triggers while looking at congruent pictures. Based on the eye-comprehension hypothesis (Carpenter and Just, 1980), this finding supports that higher cognitive load leads to longer fixations.

To be succinct, gazing behaviours, especially fixations and saccades, are studied to investigate the process of cognitive tasks such as reading, listening comprehension, and interpreting. The general findings across the studies, as reviewed above, point to the pattern that people tend to fixate longer on the items under processing when processing such items are cognitively more challenging. Tiselius and Sneed (2020) explored gaze aversion (GA) in an experimental study where participants were recruited to perform dialogue interpreting, a form of rendition where the interpreter helps two interlocutors to engage in conversations in a sentence-by-sentence manner. The authors reported increased level of GA when the interpreters were working into their L2. The increase of GA could result from several reasons. The following section is dedicated to reviewing GA and its implications.

2.3.2 Gaze Aversion (GA)

Although gazing behaviours have been found to be facilitatory in social interactions and cognitive tasks, it is not unusual for human beings to look away from others. Similar to the social and cognitive functions of engaging in gaze, Gaze Aversion (GA) has been found to bear both social (e.g., Exline, Gray, and Schutte, 1965; Modigliani, 1971; McCarthy, Lee, Itakura, and Muir, 2006, 2008; Doherty-Sneddon and Phelps, 2005) and cognitive implications (e.g., Glenberg, Schroeder, and Robertson, 1998; Doherty-Sneddon *et al.*, 2002, 2012; Doherty-Sneddon and Phelps, 2005; Bruce and Young, 2012).

2.3.2.1 Social Reasons for GA

As mentioned earlier, social norms affect human gazing behaviours. In non-contact cultures, such as Japanese culture, people are dissuaded from sustaining a mutual gaze, believing it signals rudeness (West, 1995; Morsbach, 1973; McCarthy *et al.*, 2006, 2008). Also, avoiding eye contact could be an attempt to alleviate negative feelings or self-consciousness (Doherty-Sneddon and Phelps, 2005). For example, answering embarrassing questions can reduce eye contact from the speaker to the questioner (Exline *et al.*, 1965; Modigliani, 1971). Downward gaze has been found to associate with feeling shame and humiliation (Tomkins and McCarter, 1964). Edelmann and Hampson (1979) reported that subjects' level of eye contact decreases noticeably when they feel embarrassed. Argyle and Dean (1965) find that the narrowing physical distance between interlocutors would be compensated by an increasing level of averted gaze so that the equilibrium of intimacy can be maintained. For example, they reported that when physical distance among interlocutors is fixed, if the level of intimacy exceeds the acceptable limit, gaze aversion will occur to alleviate the awkwardness. Stanley and Martin (1968) suggested that eye contact or averted gaze is affected by the level of anxiety and that one would look away to be less anxious.

2.3.2.2 Cognitive Reasons for GA

In addition to social benefits, averting gaze can also be cognitively advantageous (Bruce and Young, 2012). Glenburg *et al.* (1998) conducted a series of experiments where subjects answered general knowledge, biographical, mathematical questions. Subjects' eye movements were monitored. The experiment found that subjects were more likely to look away when answering moderately difficult questions than easy ones, and the frequency of looking away is correlated with the difficulty level of the questions. Additionally, in one of the experiments, the condition was manipulated by asking the subjects to either close their eyes or look at the questioner's nose. The performance data showed that better accuracy is achieved when subjects look away or close their eyes. Thus, Glenburg *et al*. (1998) concluded that people would avert their gaze to cope with increasing cognitive demand, and the behaviour can enhance performance. Glenburg and his colleagues further hypothesised that GA occurs at critical moments and helps people allocate more cognitive resources to the task and not be distracted by environmental stimulation (such as human faces) and thereby enhance performance – the Cognitive Load Hypothesis (Glenburg *et al*., 1998).

The Cognitive Load Hypothesis points to the notion that human faces can be cognitively demanding. As mentioned earlier, human faces attract attention more than other objects (Bruce and Young, 2012), and human beings readily detect and process signals from faces (Hershler and Hochstein, 2005). Given the rich information that can be gathered from the face and our ready tendency to attend to it, human faces are "attention-grabbing" (Bruce and Young, 2012). Russell and Lavie (2011) reported that human faces would compete with other objects for observers' visual attention, even when detecting face is irrelevant to the task.

Since human faces are attention-demandingly and informative, decoding the signals would inevitably induce a certain cognitive load level (Doherty-Sneddon *et al*., 2002). For instance, a visuospatial overload is reported in children who listen to descriptions of abstract shapes while looking at a face (Doherty-Sneddon, Bonner, and Bruce, 2001). Doherty-Sneddon and her colleagues (2002) extended Glenburg's findings by conducting an experiment where children aged five to eight answered verbal reasoning and mathematical questions of varying difficulties. They reported that 8-year-olds would certainly have acquired adult-like GA patterns in response to cognitive challenges.

Doherty-Sneddon and Phelps (2005) extended the previous findings by comparing GA in face-to-face and video-mediated communication. In the study, a group of children aged eight answered autobiographical, verbal reasoning, mathematical, and episodic memory questions of varying difficulties. The questions were asked either face-to-face or via a video link. The results showed that subjects would avert gaze from both human and video-mediated questioner, but more GAs were observed during face-to-face questioning than in the video-mediated setting. Also, the difficulty level of the questions asked played a crucial role in triggering GA, with challenging questions contributing more GAs than easy ones. The study's findings endorse the Cognitive Load Hypothesis and indicate that GA functions as a strategy to switch off from external environmental stimulation while channelling more resources to the processing of other internal information (Doherty-Sneddon and Phelps, 2005). This shift can facilitate performance for the task at hand (Phelps, Doherty-Sneddon, and Warnock, 2006).

Part II: Understanding Active Listening and Notetaking in CI

2.6 Active Listening and Notetaking in CI

2.6.1 Overview of CI

Interpreting is an activity where "language perception, comprehension, translation and productions operations are carried out virtually in parallel" (Russell, 2005, p. 136). Given its complication and high demand for skills and cognitive resources, it is widely accepted that interpreting is cognitively demanding (Danks, 1997; Shreve, 1997; Gile, 2009; Pöchhacker, 2002, 2004, 2015; Nigro, 2015). According to Nigro, interpreting is fundamentally challenging, and interpreters will always find themselves in problematic and challenging situations regardless of their experience and expertise.

Depending on the perspectives, interpreting can be defined in varying ways. From the perspective that highlights immediacy, Pöchhacker (2004, p. 13) defined *interpreting* as "a form of translation in which a first and final rendition in another language is produced based on a one-time presentation on a source language." Looking from the angle of the working mode, Setton and Dawrant (2016) refer interpreting as a form of oral translation. From the quality standard viewpoint, ASTM International, one of the influential global standards makers, defines *interpreting* as "the process of understanding and analysing a spoken or signed message and re-expressing that message faithfully, accurately and objectively in another language, taking the cultural and social context into account" (ASTM F2089-01, 2015).

Although language conversion seems to be the most distinctive part of the job, and the extraordinary language skills are indeed crucial, they are not sufficient because apart from facilitating language exchanges, interpreters shall also communicate sense (Seleskovitch, 1978; Thiéry, 1990; Pöchhacker, 2005; Albl-Mikasa, 2008; Setton and Dawrant, 2016). Sense is described as a conscious and nonverbal construct inclusive of both linguistic meanings of the words and a cognitive addition used to process the information relevant to the sound of the speech (Seleskovitch,1978). Seleskovitch also suggested that a critical difference between translating and interpreting is that while translating is dedicated to the exchange of language, interpreting aims to grasp and render the sense of the source material instead of merely the semantics. To convey both linguistic meanings and nonlinguistic sense, the interpreter should develop a skillset with four core competencies - language proficiency, knowledge, skills, and professionalism (Seleskovitch, 1978).

There are several modes of interpreting, such as Sight Translation, a process where interpreters read the source text subvocally while delivering the target speech vocally; Simultaneous Interpreting, where interpreters listen to the speech and interpret it concurrently; and CI, the interest of this study.

CI refers to "the process of interpreting after the speaker has completed one or more ideas in the source language and pauses while the interpreter transmits that information" (Russell, 2005, p. 136). It is suggested that the origin of CI can be traced back to the 1919 Paris Peace Conference (Setton and Dawrant, 2016). CI remained the default interpreting service in the UN before superseded by Simultaneous Interpreting (SI) in the late 1920s when SI-enabling technology became available (Flerov, 2013). Today, CI is still widely used because it is more cost-effective, convenient, and more private and accurate than SI (Setton and Dawrant, 2016).

CI is conducted in segments. Depending on the length of each segment, CI can be labelled as short or long. Generally, a segment that lasts longer than two minutes is considered a long CI, and even longer ones can be used for testing CI competencies in training (Setton and Dawrant, 2016). Memory is fundamental to all cognitive tasks, and it is "undoubtedly crucial" to CI because the interpreter must hold what he or she has understood from the source speech in WM for storage and manipulation until the delivery of the target speech (Pöchhacker, 2011, p. 4). According to Pöchhacker, both WM and LTM are involved in CI, but due to the dearth of research on the topic, the exact manner of memory involvement is not yet fully understood.

To alleviate the stress on WM, interpreters, when conducting long CI, usually adopt notetaking, a technique that allows them to swiftly transform the speech into a set of notes which will be utilised in producing the target speech (Pöchhacker, 2016). However, notes are not taken to substitute the source speech but indicate the ideas, structure, and necessary details

(Pöchhacker, 2015). Seleskovitch (1975) suggested the art of CI lies in notetaking, where the interpreter grasps the essence of the meaning and jots down a word or symbol (notetaking) to represent and ultimately recall (note-reading) that meaning.

Scholars have studied notetaking as a cognitively complex task in CI and non-CI scenarios, such as taking notes from lectures. Although both CI notetaking and taking notes from lectures can facilitate recall, the two practices are different in terms of purposes. While taking notes from a lecture is aimed at academic gains or learning (DiVesta and Gray, 1972), CI notes are constructed as temporal cues to back up memory (Gillies, 2017, 2019) and to grasp the sense of the source speech (Sleskovitch, 1975, 1978; Albl-Mikasa, 2008, 2017; Pöchhacker, 2015, 2016; Chen, 2016; Gillies, 2017, 2019). Notetaking induces cognitive efforts as it must be based on speech comprehension, information selection and writing (Piolat, Olive, and Kellogg, 2005). It can be particularly demanding on the CE and other WM components when the note-taker simultaneously selects key points, notes them down, and processes the incoming new information (Piolat *et al*., 2005).

Many scholars accentuate the significance of the notes taken during CI, believing they are supplements, memory triggers, or visual cues which contribute to sound outputs. CI notes have been considered visual representations of the source speech, on which the quality of outputs largely depends (Groot, 1997; Gillies, 1997, Gile, 2009; Ito, 2007; Someya, 2017). Although CI notetaking has robust individual features (Setton and Dawrant, 2016), some scholars and professionals have offered guidelines on the layout, structure, and sense (Seleskovitch, 1975; Matyssek, 1989; Gilles, 2005; Albl-Mikasa, 2008; Dingfelder, 2015; Gillies, 2017, 2019). Despite the efforts in studying and instructing the structure and function of CI notes as a product, less attention has been dedicated to studying notetaking as a cognitively demanding component of CI.

Although Notetaking is a visible effort during CI, the process of CI consists of several tasks. To highlight the consecutive nature, CI can be seen as an activity that consists of two phases, Capture Phase and Delivery Phase (Setton and Dawrant, 2016), or source-speech comprehension and re-expression in the target language (Pöchhacker, 2011), or Listening and Notetaking Phase and Target Speech Production (Gile, 2009). In the first phase, interpreters would comprehend and analyse the input (source speech) by conducting active listening and, almost simultaneously, taking notes to transform the results of active listening into written memory cues. In the second phase, the interpreter would utilise the notes to verbally present their version of the speech in the target language while exhibiting fidelity to the speech's content and the speaker's intention.

Over the years, scholars have proposed several models for CI, and perhaps the most famous model is the Interpretive Theory of Translation proposed by Seleskovitch (1978). The model depicts interpreting (and translating) as a triangular process where the source language comprehension and the target language production are separated but anchored in sense, the conscious and nonverbal cognitive construct (Seleskovitch, 1978; Seleskovitch, 1978; Seleskovitch and Lederer, 1984; Lederer, 2010; Pöchhacker, 2011). The model explicitly states the essential components of interpreting and is reflected in other more detailed models. For instance, Russell (2000, 2002a, 2005) proposed a more zoomed-in model where interpreters would resort to "syntactic knowledge, semantic knowledge, associated knowledge and background experiences, cultural awareness and contextual knowledge" while taking register and style into consideration to make sense out of the source speech (Russell, 2005, p. 145).

Some models highlight the role of memory in CI. Kirchhoff (1979) advocates a parallel storage model of CI where interpreting is enabled by two distinctive and yet mutually complementary storage systems, material storage aided by notetaking and memory storage facilitated by cognition. This parallel storage model is reflected in the Effort Model proposed based on the hypothesis that interpreting taxes WM, a limited capacity, and interpreting would overburden WM and result in performance deterioration (Gile, 2009). According to the Effort Model (Gile, 2009), CI is performed in two phases, consisting of several tasks. In the first phase, where active listening and notetaking are conducted, interpreters need to perform listening and analysis, notetaking, short-term memory operations, and coordination concurrently. In the second phase, remembering, note-reading, and production are needed. The Effort Model implies that in each phase, the combined mental effort of all tasks must not exceed the total capacity of WM to avoid performance deterioration. The Effort Model for CI made it explicit that interpreting is hinged on WM and that WM overload can hinder performance, which can be inevitable. However, it remains opaque what could lead to WM overload and the reason for the causal relationship.

To be succinct, CI, as a significant mode of interpreting, is a cognitively complex process where the interpreter shall facilitate the communication of semantics between two or more languages and convey the sense of the speech by resorting to linguistic and nonlinguistic knowledge. Performed in two stages, CI involves active listening, speech analysis, and notetaking in the first stage and note-reading, recall, and producing target speech in the second. CI requires multi-tasking and is memory-dependent, especially in the first stage when constantly incoming information must be held in STM, queuing to be written as notes. To circumvent WM overload and refrain from failing, interpreters generally adopt notetaking which consumes WM resources and imposes a certain level of cognitive load. Considered as memory cues and visual representations of the source speech, CI notes rely on successful active listening and accurate speech analysis.

2.6.2 Listening in CI

2.6.2.1 Overview of Listening Comprehension

Listening comprehension is one of the most frequently used techniques for communication and language usage, taking up 45% of the total time spent on communication, followed by speaking (30%), reading (16%), and writing (9%) (Lee and Hatesohl, 1983). However, the popularity by no means implies simplicity. Listening comprehension is considered a highly complex task (Richards, 1983; Cutler and Clifton, 1999; Buck, 2001; Vandergrift, 2004) and is described as "complex, dynamic, and fragile" (Celce-Murcia, 1995a, p. 366). Processing auditory input prompts someone to apply linguistic and non-linguistic knowledge to fulfil several subsidiary tasks and requires support from all aspects of the cognition system. Richards (1983) suggested an extensive range of micro-skills involved in varying types of listening comprehension. According to Richards, as many as thirty-three listening skills are relevant to conversational listening, covering retention of speech chunks, recognising linguistic and contextual features, anticipating and inferring meanings, and adjusting listening strategies. For academic listening (such as in lectures), Richards suggested a learner should be equipped with eighteen sub-skills to engage in tasks such as identifying the topic and discourse markers (e.g., conjunctions), recognising intonational signals, filtering irrelevant information (e.g., jokes), handling accents and speech rate, and inferring relationships (e.g., cause and effect). Compared with listening to one's first language (L1), comprehending a speech in a second language (L2) is less automatic and more cognitively demanding (Piolat, Barbier, and Roussey, 2008) and would expose the listener to more challenges in some respects, such as recognising uncommon words (Witzel and Forster, 2012; Cop, Keuleers, Drieghe, and Duyck, 2015; Brysbaert, Lagrou, and Stevens, 2017; Field, 2019) and sentence processing(Clahensen and Felser, 2006b; Hopp, 2015; Mitsugi and MacWhinney, 2016; Romero-Rivas, Corey, Garcia, Thierry, Martin, and Costa, 2017), regardless of the proficiency (Clahsen and Felser, 2006a). This disadvantage of L2 processing implies that interpreting a speech uttered in L2 is arguably more cognitively expensive than that in L1.

It is generally acknowledged that listening comprehension as a process can be top-down and bottom-up (Anderson and Lynch, 1988; Celce-Murcia, 1995a; Tsui and Fullilove, 1998; Cutler and Clifton, 1999; Buck, 2001; Vandergrift, 2004, 2011; Yeldham, 2018; Field, 2019; Díaz-Galaz, 2020; Sekkal, 2020). Top-down processing, or macro-processing, describes the process where the listener constructs a conceptual representation of the speech by processing its initial linguistic or contextual cues via schematic and contextual knowledge (Marianne Celce-Murcia, 1995a; Tsui and Fullilove, 1998). According to Carrell and Eisterhold (1983), schematic knowledge consists of content and formal schemata. Content schemata refer to background knowledge about the speech, such as the thematic information. On the other hand, formal schemata refer to the knowledge about the speech structure, genre, rhetoric, and formality. Contextual knowledge refers to the awareness of the situation of the speech, such as participants and the speaker's intention, and the understanding of the speech (Marianne Celce-Murcia, 1995b). Vandergrift (2011) suggested that top-down processing employs experiential, cultural, textual, linguistic, or pragmatical knowledge.

The bottom-up perspective views listening comprehension as a one-way process consisting of several consecutive stages, resulting in a lower stage becoming the input for the next higher stage (Buck, 2001). The main feature of the bottom-up approach is accretion – the listener constructs meanings of the speech in a gradual and fixed order starting from the phoneme level to the final discourse level (Vandergrift, 2004). Cutler and Clifton (1999) depicted a blueprint for bottom-up listening where listening comprehension is built up from filtering the speech from the background noise to forming a representation on a discourse level, a process where the listener needs to adopt linguistic and non-linguistic skills incrementally and comprehensively, which can stretch one's WM capacity.

Buck (2001) advocates for top-down processing. According to Buck, the bottom-up approach possesses a significant flaw - processing different types of knowledge does not necessarily happen in a fixed sequence. Instead, the varying types of knowledge can join in the listening process in any order that is deemed convenient. From his point of view, recognising a word does not necessarily require the listener to decode the acoustic information as it can be achieved by utilising syntactic information. Buck establishes his proposal on the rationale that a listener would almost always have expectations for the upcoming utterance as she or he has knowledge of the world and the situation where listening comprehension occurs.

Comprehension of the input would then result from either confirming or rejecting the expectations as the speech unfolds. For example, when bidding farewell to friends, one would understand the words not by decoding the acoustic signals but by paying attention to the situation where people are waving or walking away.

Other scholars could refute Buck's proposal as they believe that top-down and bottom-up approaches represent different cognitive processes and that the two dynamics co-exist. By propelling the listener to resort to schematic knowledge, top-down processing is facilitated by LTM and therefore does not take up much WM resources (Tyler, 2001). However, schemata do not always suffice to contribute to successful listening comprehension - they can be insufficient or incompatible with the material. For example, Field (2003) suggested that schematic knowledge would facilitate the recognition of "*want*" or "*won't*" in written text, but phonetic interference arising from similar pronunciations could make it difficult to tell the two words apart during listening. Psycholinguists tend to adopt the opinion that top-down processing is practical but only to a degree and that bottom-up processing shall be adopted when schematic knowledge falls short (Celce-Murcia, 1995a; Tsui and Fullilove, 1998; Tyler, 2001; Hulstijn, 2001). For example, Tyler (2001) suggested that top-down processing is suitable for situations where quick acquisitions are needed (e.g., for tourists). Hulstijjn (2001) emphasises that a listener will not obtain a meaningful rendition of all aspects of acoustic cues unless bottom-up processing is adopted. Tsui and Fullilove (1998) find that less-skilled L2 users typically resort to the top-down approach even when schematic knowledge is incongruent with the text. By contrast, skilled L2 users would switch to bottom-up processing and build up the rendition when schematic knowledge becomes insufficient. According to Tsui and Fullilove, the reason for the difference is that less-skilled L2 users have not yet acquired the dexterity to perform bottom-up processing, indicating that bottom-up listening is a robust discriminator for language proficiency. Similarly, Celce-Murcia (1995b) proposed that while bottom-up processing is automatic to native and skilled L2 speakers, it is less natural and can damage comprehension for less-skilled L2 users.

It shall be noted that, despite the difference in top-down and bottom-up processing, successful listening comprehensions are rarely achieved by only using one of the two approaches (Vandergrift, 2004, 2011). Vandergrift (2011) suggested that successful listening comprehension is gained via a judicious interaction between the two dynamics, indicating an informed switch from or to either approach and that this switch can be triggered by factors such as the purpose of listening, language proficiency, and the context where listening comprehension is performed. According to Vandergrift (2011), while top-down processing would suffice to obtain an overall picture of the speech, bottom-up listening should be enlisted if the speech is processed in detail. Celce-Murcia (1995b) suggested that top-down and bottom-up processing operate simultaneously towards compression and intersect at the discourse level where background and contextual information is combined with linguistic information to propel interpretation.

2.6.2.2 Active Listening in CI

As a standard behaviour that is practised constantly, listening has varying types. Sweller *et al.* (2011) divided knowledge and skills into biologically primary and secondary, with the former referring to the abilities that human beings are born with or evolve to learn. Listening, therefore, would arguably have a biologically primary nature - hearing without requiring skills. For example, the author can hear the noise from typing this sentence without enlisting any noticeable effort. This aware-yet-detached listening mode is very similar to what Setton and Dawrant (2016) classify as *passive listening*, where someone notices the sound without processing it. From the perspective of human information processing (see 2.1.2), passive listening would indicate that the stimuli are captured by the sensory register but not attended to and has proved futile in facilitating recall (Baddeley, 2004).

Setton and Dawrant (2016) also suggested another two classifications for daily listening, *superficial listenin*g and *selective listening*. Both superficial and selective listening demand attention, but the former results in a superficial understanding of only the surface features of a speech as the attention invested is not enough to reach in-depth comprehension, whereas the latter allows the listener to cherry-pick useful information towards a specific task such as succeeding language proficiency tests. According to Setton and Dawrant, although selective listening is the most attentionally demanding and requires the most intensive mental processing among the three modes of listening, it still cannot engender a thorough understanding of the discourse as it sifts germane information and neglects the less relevant. According to Gillies (2019), listening comprehension skills that cater to language proficiency testing is inadequate for CI, and interpreters should perform *active listening* to attain adequate comprehension.

Scholars in interpreting studies have reached a consensus that active listening bears significance. For example, Setton and Dawrant deemed it the most significant skill for a faithful rendition of the source speech's semantic meaning and contextual sense, believing the likelihood of successful interpreting climbs as the interpreter invests more effort in active listening. According to Setton and Dawrant, successful active listening enables interpreters to construct a "structured, coherent mental model with a rich encoding of speaker meaning, details and desired communicative effects" (p. 87). Herrero (2017) voiced a similar notion that the successful execution of interpreting relies on active listening. In terms of interpreting training, Jin (2019) proposed that grasping the difference between passive and active listening forms the first step for trainee interpreters. Compared with passive and selective listening, active listening requires the listener to commit unmitigated concentration as it encompasses "all comprehension-oriented operations, from the subconscious analysis of the sound waves carrying the source-language speech which reach the interpreter's ears through the identification of words to the final decisions about the 'meaning' of the utterance" (Gile, 2009, p. 160). Zhang and Wu (2017) offered a similar description of actions involved in active listening, including word recognition, information segmentation, idea progression, logic apprehension, and structure representation.

A key reason to explain the necessity of active listening would arguably lies in interpreters' pursuit of faithfulness, an aspect deemed by scholars in interpreting studies as the nucleus in assessing the quality of interpreting (Macías, 2006; Bartłomiejczyk, 2007; Hale, 2007; Setton and Motta; 2007; Ke and Zhang, 2008; Lee, 2008; Gile, 2009; Choi, 2013; Wu, 2013; Pöchhacker 2002, 2015). The word "faithfulness" is interchangeable with other terminologies such as fidelity or accuracy and completeness (Hale, 2007; Pöchhacker, 2015). Despite the varying terms, they underscore the necessity of producing target speech in agreement with the source text regarding semantics and sense. The goal of producing a target speech that truthfully communicates the meaning and sense implies that the interpreter, instead of listening to the speaker selectively, should perform active listening by "maintain [ing] a high level of concentration and depth of processing over the entirety of the discourse" (Setton and Dawrant, 2016, p. 84), attending to the change of speaker's intonation during active listening because such phonetic variance could signal word boundaries (Gillies, 2019), and analysing the structure and logic of the speech and deducing the speaker's intentions (Gile, 2009).

The pursuit of faithfulness, apart from comprehensive and accurate processing of linguistic and semantic information, also demands the interpreter to avoid word-for-word translation but to pursue sense consistency with the source text (Seleskovitch, 1978; Kurz, 1993; Pöchhacker, 2015; Wang, 2016; Setton and Dawrant, 2016). Seleskovitch defines sense as:

> A cognitive construction made by the addressee based on the sounds he received from the addresser's mouth; he adds to them such cognitive remembrance as fits the sounds, and such additional knowledge, whether from his long or medium memory, that fits the whole of a clause or sentence (Seleskovitch, 1978, p. 336).

According to Selekovitch, interpreters shall make efforts at and beyond the linguistic level to make sense of the source speech. Linguistically, the audience is often ignorant of the source language, meaning that interpreting the mere word content of the source speech does not suffice. Instead, it is essential for the interpreter to convey the ideas concealed in the original speech (Herrero, 2017).

From the non-linguistic perspective, to make sense is primarily to fulfil the purpose of communication (Selekovitch, 1978; Setton and Dawrant, 2016). According to Selekovitch, sense is attained when the language is used to communicate. Specifically, she explains that people using different languages do not use identical linguistic constructions to convey the same ideas, meaning that word-for-word interpretations would baffle the audience as they are confined in the source language grammars. Therefore, the discrepancy in language using habits demands the interpreter to break through the linguistic ceiling of the source language and deliver the ideas that the speaker aspires to communicate. Similar ideas are suggested by Setton and Dawrant, who pointed out that by delivering the speech, the speaker wishes to both explicit and implicit information. According to the authors, while rendering the explicit information can be semantically straightforward, the implicit information, or the speaker's intents, can only be faithfully inferred when listened empathetically, requiring the interpreter to stay attentive to the connotations of the speaker. Similarly, Herron suggested that interpreters shall stay vigilant to both verbal and non-verbal information projected by the speaker and listen empathetically to the interpreter between the lines, underscoring the significance of active listening.

Based on the above literature, it can be deduced that the successful execution of active listening would require interpreters to resort to both bottom-up and top-down processing, with the former compelling interpreters to gradually construct meanings from the smallest units towards the discourse level (Cutler and Clifton, 1999; Buck, 2001) and the latter requring interpreters to apply schematic and contextual knowledge stored in LTM to form a conceptual representation of the speech at the discourse level (Marianne Celce-Murcia, 1995a; Tsui and

Fullilove, 1998; Buck, 2001; Vandergrift, 2011). Although the two processing modes differ in direction, both require interpreters' unmitigated attention to achieve accurate comprehension, and ultimately, faithfulness in the target speech.

2.6.2.3 Process in Active Listening

Despite the gravity of active listening, there has been a dearth of studies on the subject, and therefore, the exact process of active listening remains unclear. However, by comparing the literature, active listening can arguably be conceptualised as a process where the interpreter conducts bottom-up and top-down listening with unmitigated concentration. To illustrate the cognitive demands of active listening and fully understand active listening as a cognitively challenging task, the following section, based on the blueprint for bottom-up listening comprehension (Cutler and Clifton, 1999) and the literature on top-down processing (Marianne Celce-Murcia, 1995a; Tsui and Fullilove, 1998; Buck, 2001; Vandergrift, 2011), illustrates active listening in CI as a four-stage process. However, as suggested by Buck, the process might not take place in a fixed order. Instead, the varying skills can be applied in any order that is deemed convenient. Therefore, the following review of the four stages of active listening is mainly geared towards demonstrating the subtasks that interpreters would need to complete during active listening.

Cutler and Clifton (1999) depicted a blueprint of listeners, where the bottom-up process of listening unravelled into several mental operations containing subtasks. According to Cutler and Clifton, listening begins with a decoding stage, followed by segmenting speech into components, processing words and utterances, and integrating auditory cues into a discourse. Each of the four stages is reviewed in detail as follow:

2.6.2.3.1 Decoding

According to Cutler and Clifton (1999), decoding is necessary for comprehending speech. Speech can be seen as a collection of codes in varying forms, and decoding a speech begins with isolating speech-related signals from the noise in the environment (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). A listener would then decode the input into phonetic segments, the smallest units that can be utilised to tell two spoken words apart. For instance, "leaf" and "leave" share the same first (/l/) and second /i:/ phoneme but differ on the third phoneme, /f/ and /v/, respectively. To avoid confusing "leaf" with "leave", the listener needs to draw upon her or his knowledge on the articulation of the two consonants.

2.6.2.3.2 Segmentation

Cutler and Clifton (1999) suggested that a necessary step towards understanding a speech is to dismantle it into words. Unlike written language, where words are segmented by space or punctuation marks, speech is usually delivered in a continuum, meaning that the boundary between two words are invisible or blurry, which is a distinctive characteristic of spoken language (Cutler and Clifton, 1999; Buck, 2001; White, 2018). The boundaries between words must be located to propel the bottom-up processing to the next level.

2.6.2.3.3 Recognition

The third stage of active listening process is word recognition based on acoustic, syntactic and thematic information from the speech. As reviewed in 2.1.2.1.3, knowledge about words and their meanings is stored in a part of LTM: the semantic memory (Tulving, 1972; Baddeley, 2009). Cognitive psychologists regard this storage as the mental lexicon and agree that word recognition initiates from searching the lexicon for a target word in terms of its form and meaning (McClelland and Elman, 1986; Norris, 1994; McQueen and Cutler, 1998; Cutler and Clifton, 1999; Vitevitch *et al*., 2018). In English, it is generally accepted that the language has 44 phonemes, among which 24 are consonants, and 20 are vowels (Malah and Rashid, 2015; Bizzocchi, 2017). The discrepancy between the number of phonemes and words in one's lexicon implies that words are not always distinctive from each other (Cutler and Clifton, 1999; Cutler, Norris, and Sebastián-Gallés, 2004), implying that setting two words apart requires mental efforts, and therefore, induces cognitive load.

Psycholinguistics tend to agree that word recognition from speech is an interactive and competitive process where an utterance, upon its onset, would activate a repertoire of candidate words (also known as "neighbours") that compete against each other as more information arrives (Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1987; Grainger, 1990; McQueen and Cutler, 1992; Norris, 1994; Cutler and Clifton, 1999; Otake and Cutler, 2013; Grosjean, 2018). According to the authors, the number of the activated words would decrease until a winning word has triumphed. The concurrent activation is depicted in the Cohort Model proposed by Marslen-Wilson and Tyler in 1980. According to the Cohort Model, upon hearing the initial phoneme, an initial cohort containing all words sharing the same initial acoustic feature will be activated from the lexicon. The initial cohort would shrink due to the increasing input of acoustic information. Ultimately, a recognition moment will occur as the cohort is reduced to containing only a single target word. Other models, such as TRACE (McClelland and Elman, 1986) and Shortlist (Norris, 1994), inherit the feature of concurrent activation from the Cohort model while highlighting competition among candidates. For example, McClelland and Elman include in the TRACE model the lateral inhibition between competitors on each level (auditory features, phonemes, and words) so that only one winning output is selected at the word level.

2.6.2.3.3 Integrating to Discourse and Applying Top-Down Listening Skills

The last component of the bottom-up listening comprehension process, according to Cutler and Cliffton (1999), is the stage where words are jointed into sentences that must be comprehended semantically and pragmatically at a discourse level. It is also the stage where listeners process the utterance in a top-down fashion by utilising a series of schematic and contextual knowledge (Marianne Celce-Murcia, 1995a; Tsui and Fullilove, 1998; Buck, 2001; Vandergrift, 2011).

According to Friederici (2002), sentence processing consists of three phases. The first phase describes the windows where a listener conducts initial syntactic analysis by determining the syntactic nature of each word (e.g., a noun or a verb) and processing morphosyntactic information (e.g., singular or plural, subject or object). The second phase concerns processing words' meanings and combining them into a representation regarding lexical and pragmatic aspects (Hartsuikker, 2018). Thematic roles are assigned to each word in the second stage as the sentence unfolds (e.g., who does what to whom). The last phase is a revision window where semantic and syntactic information is re-analysed.

A body of literature supports the notion that speech comprehension on the discourse level follows the principles of interactivity, incrementality, and prediction (Altmann and Steedman, 1988; Tanenhaus, Spvey-Knowlton, Eberhard, and Sedivy, 1995; Traxler and Pickering, 1996; Allopenna, Magnuson, and Tanenhaus, 1998; Altmann and Kamide, 1999; Staub and Clifton, 2006; Ferreira and Lowder, 2016; Brennan and Pylkkänen, 2017; Hartsuiker, 2018; Pickering and Gambi, 2018; Rubio-Fernandez and Jara-Ettinger, 2020). An excellent example of the interactivity principle would be Tanenhaus *et al.'s* (1995) experiment, where an eye-tracker was used to observe subjects' eye-movements while participants processed auditory instructions and looked at several visual references. The results showed that subjects would instantly make saccadic eye movements to the relevant visual reference after hearing corresponding words, suggesting interactivity between spoken language processing and visual input. In the same study, the authors observed subjects' eye-movements during processing the syntactically ambiguous sentence "put the apple on the towel in the box". Eye-tracking data suggested that although subjects would mistake "on the towel" as the destination for the apple, they would commit to the correct interpretation ("box" as the destination) as the sentence unfolds, suggesting that processing options narrow down as the audio inputs increment.

Staub and Clifton (2006) explored the prediction principle by using "*Either…or…"* sentences. They reported that perceiving "either" would prompt the subjects to anticipate the upcoming input to include "or". Some studies proposed that reliable predictions are likely based on one's language production system, a mechanism coined as prediction-by-production by Pickering and Gambi (2018), or P-chain by (Dell and Chang, 2014). According to Pickering and Gambi (2018), to predict the upcoming speech, a listener needs to firstly engage in incremental processing to comprehend the speech before covertly imitating what has been uttered. Non-linguistic information (e.g., background knowledge) is processed to deduce the speaker's intention, which accords with the upcoming input. Finally, the listener would need to run the deduced intention in her or his corresponding language production system in anticipation of the upcoming utterance. They also indicated that prediction-by-production as a prediction mechanism is optional and does not constitute a prerequisite for successful comprehension.

In short, CI active listening as a process encompasses several stages where bottom-up and top-down skills are required to attain faithful comprehension of the source speech. To process the speech in a bottom-up manner, interpreters are compelled to utilise a series of skills including decoding the input into phonetic segments, establishing word boundaries, recognising words, and finally integrating processed words into sentences, a stage where top-down skills, i.e., schematic and contextual knowledge.

2.6.2.4 Cognitive Efforts During Active Listening

As discussed in 2.6.2.1, listening comprehension is considered as a highly complex process that requires a listener to utilise linguistic and non-linguistic knowledge to fulfil several subsidiary tasks and demands comprehensive supports from one’s cognitive system

(Richards, 1983; Celce-Murcia, 1995b; Cutler and Clifton, 1999; Buck, 2001; Vandergrift, 2004). Among several modes of listening, active listening bears pivotal significance in achieving high standard CI performance (Herrero, 2017) and requires interpreters to commit unmitigated concentration throughout the process to deduce not only the linguistic and semantic meanings but also the sense of a speech (Seleskovitch, 1978; Gile, 2009; Pöchhacker, 2015; Setton and Dawrant, 2016). To succeed in active listening, interpreters need to build up coherence mainly through bottom-up processing during which they infer the meaning of the source speech by following the guidance of the incremental linguistic inputs (Albl-Mikasa, 2017). Similarly, according to Cutler and Clifton's (1999) blueprint for listening, bottom-up processing is underpinned by activation of a series of linguistic knowledge ranging from the most basic acoustic rhythm to the building of coherence. To recap, the bottom-up processing of active listening propels interpreters to gradually build up the coherence by firstly decoding the speech into phonemes before locating boundaries between words. The interpreter would then endeavour to recognise each word before entering the final stage where all the information collected from the previous three stages are integrated towards comprehending sentences and, ultimately, the discourse.

In a broad sense, performing active listening would impose cognitive load as performing the task requires cognitive support. Previous review (2.6.2.3) has highlighted the utilisation of linguistic, semantic, and pragmatic knowledge during listening comprehension. According to Kintsch (1998), one must access LTM to apply knowledge during comprehension. Based on the construct of human memory, factual information, such as linguistic knowledge, is stored in long-term declarative memory, a store also known as explicit memory because accessing the store demand explicit efforts in choosing strategies (Shiffrin and Atkinson, 1969; Graft and Schacter, 1985; Squire, 1992; Baddeley, 2009; Vakil, Wasserman, and Tibon, 2018; Wang, 2020). Voluntary retrieval from LTM is also susceptible to interreferences (Fernandes and Moscovitch, 2000), and bypassing unwanted responses requires mental efforts (Jacoby, Woloshyn, and Kelly, 1989; Kane and Engle, 2000). Another aspect that indicates the cognitive costs of LTM retrieval lies in the advancement in researching the role of WM (Baddeley, 2000, 2009; Schelble, Therriault, and Miller, 2012; Unsworth, Brewer, and Spillers, 2013). One of the implications yielded from developing WM is that WM interacts with LTM during cognitive tasks, which propelled the proposal of EB, Episodic Buffer, by Baddeley (2000). As discussed in 2.1.2.3.4, the buffer allows varying forms of input from WM stores and LTM to gather and interact, by which WM is linked to LTM (Baddeley, 2000). EB is believed to be a limited capacity and can hold about four episodes (Baddeley, 2000; Cowan, 2005), highlighting the possibility of EB overload. The significance of WM in LTM retrieval is amplified by the myriad of studies that has accentuated the role of WM capacity in LTM retrieval by suggesting that individuals with higher WM capacity tend to outperform those with lower WM capacity (e.g., Kane and Engle, 2000; Baddeley, 2000, 2009; Schelble *et al*., 2012; Unsworth *et al*., 2013). For example, Unsworth *et al*. (2013) reported that high-capacity individuals are more able than their counterparts in selecting and utilising appropriate retrieval strategies when accessing LTM. However, Scheble *et al*., (2012) suggested that the high-capacity advantage can diminish due to high cognitive load. Collectively, the findings indicate that LTM retrieval is underpinned by WM resources, and therefore carries cognitive load.

From a zoomed-in perspective, active listening process consists of several sub-cognitive tasks. According to Cutler and Clifton (1999), the bottom-up processing of active listening propels interpreters to gradually build up the coherence by firstly decoding the speech into phonemes before locating boundaries between words. The interpreter would then endeavour to recognise each word before entering the final stage where all the information collected from the previous three stages are integrated towards comprehending sentences and, ultimately, the discourse. To decode the speech, one must isolate speech signals from the environment. Studies on speech isolation have shown that picking up speech from the background noise requires cognitive control (Heinrich, Schneider and Craik, 2008; Anderson and Kraus, 2010). According to Heinrich *et al*., (2008), background noise can hinder memorising of uttered information by preventing the speech from being adequately processed, and to segregate speech from interference, the listener would need to expend WM resources that otherwise would be dedicated for encoding. Anderson and Kraus (2010) reported that speech processing requires the perceiver to segregate speech signals from the environment and hold it in WM while spending WM resources in suppressing irrelevant noise. Decoding the sifted information would rely on accessing the declarative memory where linguistic knowledge is stored (Tulving, 1972), meaning that more efforts must be spent as controlled LTM retrieval can be resource intensive (Jacoby *et al*., 1989; Kane and Engle, 2000). Given the limited capacity of WM, the interpreter's cognitive system can be further burdened when the number of information that need to be held in his/her WM system exceeds the capacity.

Locating word boundaries also imposes cognitive load as it requires interpreters to mentally process a range of sub-lexical cues (such as stress) and lexical cues (such as semantics, syntax, and pragmatics) (Sanders and Neville, 2000; Mattys, White, and Melhorn, 2005;

Mattys, Melhorn, and White, 2007). For English, the most basic and frequently utilised cue is stress among the varying cues, the most basic and universal one is postulated to be stress (Cutler and Clifton, 1999). Studies reported that English speakers excel at exploiting stress (strong syllables) to detect word boundaries (e.g., Norris and Cutler, 1988; Cutler and Clifton, 1999), and such habit is believed to be adopted by L2 users of English (Cutler, Mehler, Norris, and Segui, 1992; Sanders, Neville, and Woldorff, 2002; Tyler and Cutler, 2009; Lin and Wang, 2018). Apart from stress, studies have found that listeners also resort to other signals to detect word boundaries in English. Sanders and Neville (2000) investigated the involvement of lexical, syntactic, and stress cues in speech segmentation, and suggested that compared with only taking advantage of one cue, utilising multiple cues better contribute to accurate detection of word boundaries. Mattys *et al.*, (2005) proposed a hierarchical framework to point out that listeners do not assign the same weight to all speech segmentation cues. Specifically, listeners prefer vocabulary-based lexical information over signal-based sub-lexical cues (e.g., rhythmic features), although the former would arguably be more cognitively stressful than the latter.

The efficiency of using word recognition as a strategy to segment speech is also documented in other studies (Marslen-Wilson and Welsh, 1978; Dahan and Brent, 1999; Norris and Cutler, 1995; Norris, McQueen, Cutler, and Butterfield, 1997; Sanders and Neville, 2000). For example, it is suggested that partial recognition of a word can benefit listeners in accurately predicting the rest of the word and thus determining its boundary to the subsequent adjacent word (Marslen-Wilson and Welsh, 1978; Sanders and Neville, 2000). Dahan and Brent (1999) proposed that listeners favour segmenting utterances into familiar words or word-like units to mitigate the burden from processing incoming material. Although lexical cues are favoured and more efficient, recognising words, the third stage in the active listening model, can be an arduous process.

Studies have reported that audio word recognition is susceptible to several challenges. A myriad of work has demonstrated that word recognition is sensitive to phonological interferences (e.g., Rubenstein, Lewis, and Rubenstein, 1971; Pexman, Lupker, and Jared, 2001; Lagrou, Hartsuiker, and Duyck, 2011; Otake and Cutler, 2013; Rose, Spalek, and Rahman R, 2015). As previously mentioned, word recognition during speech comprehension entails a memory search process where candidate words would interact and compete until the target word is located (Marslen-Wilson and Tyler, 1999; Otake and Cutler, 2013; Grosjean, 2018). One of the implications of such interactive and competitive nature is that words with scanty neighbourhoods are recognised faster and more accurately than words with dense neighbourhoods, indicating that fewer competitors contribute to faster isolation (Luce and Pisoni, 1998; Vitevitch, Stamer, and Sereno, 2008). For example, homophones (words that are different in meaning but identical in pronunciations) are found to prolong reaction times and induce high error rate during word recognition tasks (Rubenstein *et al*., 1971; Pexman *et al*., 2001; Lagrou *et al*., 2011; Otake and Cutler, 2013). Similarly, semantic ambiguity (homonymy and polysemy) is believed to interfere with lexical decision processes which is based on feedback activation from the semantic memory (Balota, Ferraro, and Connor, 1991; Hino and Lupker, 1996; Gottlob, Goldinger, Stonem and Van Orden, 1999; Piercy and Joordens, 2000; Hino, Lupker, and Pexman, 2002; Pexman and Hino, 2004; Hoffman and Woollams, 2015). In specific, Words with higher semantic diversities (e.g., story, lean) are demonstrated to facilitate the identification of a word (Piercy and Joordens, 2000; Hino, Pexman, and Lupker, 2006; Hoffman and Woollams, 2015) but prolongs the time required to assign appropriate meanings (Piercey and Joordens, 2000). According to Hino *et al*. (2002), semantically ambiguous words, owing to their multiple meanings, have more entries in the mental lexicon than an unambiguous word (e.g., food), and therefore, would induce a "many-to-one" mapping in the sense that many semantic meanings contribute to searching for one word (p. 707). However, assigning meanings to semantically ambiguous is likely to be frustrated by "one-to-many" (p. 688) situation where fuzzy representations of one word are tricky to compare.

Pioneered by Preston (1935), word frequency has long been accepted by scholars as a powerful element that influences word recognition in the fashion that high-frequency words (HFWs) are recalled and processed more quickly and effectively than low-frequency words (LFWs), a phenomenon known as the Word Frequency Effect (WFE) (Forster, 1976, 1981; Scarborough, Cortesem and Scarborough, 1977; Mutter and Hashtroudi, 1987; Connie, Titone, and Wang, 1993; Morrison and Ellis, 1995; Keuleers, Diependaele, and Brysbaert, 2010; Stevens, 2017; Brysbaert, Mandera, and Keuleers, 2018; Vitevitch *et al*., 2018; Neville, Raaijmakers, and Maanen, 2019; Popov and Reder, 2020). The robustness of word-frequency during word recognition is manifested in its role in organising the mental lexicon. It is suggested that words in one's mental lexicon are stored in descending order with the highest-frequency words stored at the top and the lowest at the bottom (Forster, 1976, 1981). However, Aitchison (2012) suggested that mental lexicon is organised as a network instead of in a frequency-descending fashion. To explain WFE from another perspective, Glanzer and Ehrenreich (1979) suggested that words are stored in two lists in mental lexicon, with one being a ready-access list containing high frequency words, and the other being a complete list that will be resorted to when a word cannot be found on the ready-access list. Morton (1979) suggested that the reason why frequent words are processed faster is not because they are stored higher than rare words in the lexicon, but because common words have stronger mental presentation due to exposure.

Since word recognition relies on memory search, processing a LFW would induce a much higher cognitive price than dealing with a common word as the former requires more efforts in searching. Given that WM is a limited capacity (Baddeley and Hitch, 1974; Baddeley, 2009, 2000, 2011; Cowan,2009) and that the source speech in CI could last up to 20 minutes (Diriker, 2015), dedicating too much time for a LFW could mean the rapidly updated new verbal inputs squeezed in the limited space and cannot be promptly processed, which would lead to cognitive overload. However, according to Brysbaert, Mandera, and Keuleers (2018), it is essential to note that not all rare words are tricky to process as some of the LFWs are compounded, derived, or inflected from common words (e.g., “supersensitive”. Also, Brysbaert *et al*. suggested that some words, albeit their low frequencies, are easy to remember for some people. For example, whereas “Voldemort” and “Dumbledore” would challenge some as they are low in frequency, they could be recognised easily by Harry Potter fans. A significant aspect of WFE is that the power of low rare words would fade due to exposure. However, interestingly, studies showed that the disadvantage for LFWs is reversed in tasks that require the individual to retrieve information from episodic memory, such as telling whether a word is old or new, suggesting that LFWs would pose a stand-out effect as new-comers (Guttentag and Carroll, 1997; Malmberg, Steyvers, Stephens, and Shiffrin, 2002; Yonelinas, 2002; Neville, *et al*., 2019). Based on rare words' advantage in recognition and disadvantage in meaning retrieving, it can be deduced that, during listening, while detecting a rare word might be swift, recalling its meaning can be cognitively arduous.

Syntactic information can also impinge word recognition process (Goodman, McClelland, and Gibbs, 1981; Sanz-Torrent, Andreu, Ferreiro, Coll-Florit, and Trueswell, 2017). Sanz-Torrent *et al*. reported a positive correlation between the number of arguments that a verb takes and the response latency. Specifically, intransitive verbs are processed faster than ditransitive verbs, with transitive verbs falling between.

According to the above authors, verbs that take more arguments induces a higher level of representational complexity and can be associated with more uncertainties, and, therefore, would require more processing capability and prolong lexical decisions. Although syntax has been found to affect listening comprehension so far, its strength in increasing cognitive costs is more forceful in the final stage of active listening.

The final step of active listening, according to the blueprint, is the stage where all results from previous processing are jointed into sentences (Cutler and Clifton, 1999). It has been suggested that the integration of verbal information during speech comprehension is very rapid and without conscious efforts (Osterhout, 1994; Ferreira, Baily, and Ferraro, 2002; Kamid, Scheepers, Attmann, 2003). However, comprehending verbal sentences is found to be resource intensive as it requires listeners to not only assign lexical and semantic meanings to words, but also mentally compute the syntax of the sentences (Chomsky, 1996; Caplan, Alpert, and Waters, 1998; Wingfield and Tun, 2007; Seeber, 2011; Gordon and Lowder, 2012; Frazier, 2013). Syntax, according to scholars, are rules that dictate the propositional aspects of sentence meanings (Caplan and Waters, 2001) and determine the relations among components (Seeber, 2011). Scholars have reported that syntactic complexity is found to affect the load on WM capacity significantly (e.g., Miyake, Carpenter, and Just, 1994; Wingfield and Tun, 2001). According to Miyake *et al.* (1994), increased syntactic complexity could overburden WM, which would result in slowed-down processing (longer RT) and inaccurate comprehension. Wingfield, Pelle, and Grossman (2003) further suggested that the disadvantage of syntactically complex sentences would sustain regardless of sentence length, underlining the strong influence of syntactic complexity in language comprehension.

A wealth of studies on syntactic complexity during language comprehension has shown a sustained focus in comparing object-relative clauses with subject-relative clauses (e.g., Just and Carpenter, 1992; Ferreira *et al.*, 1996; Gibson, 1998, 2000; Fedorenko *et al.*, 2004; Fallon *et al.*, 2006). These studies suggested that objective-relative clauses impose higher load on WM than subject-relative clauses because the former requires the listener to hold more WM information before integration occurs. Using BOLD fMRI, studies (e.g., Cooke, Zurif, DeVita, Alsop, Koenig, Detre, Gee, Pinãgo, Balogh, and Grossman, 2001) showed that more blood oxygen was required in the cortex that supports sentence comprehension when subjects were faced with object-relative clauses than subject-relative clauses, suggesting that the former is more challenging.

To explain the higher cost associated with processing object-relative clauses, Gibson (2000) proposed the Dependency Locality Theory (DLT), according to which the cognitive expense during language processing is affected by the distance between the word and its obligatory syntactic components, with longer distance resulting in higher level of processing load. According to Gibson (1998, 2000), language comprehension requires memory resources to store (memory cost) and integrate (integration cost) information. Gibson (1998) suggested that there is a memory cost because making grammatical sense of a sentence requires the remembering of each necessary syntactic component, and the minimal number of the syntactic components that must be held in memory is two - a head noun (the subject) and a head verb (the predicate). A significant implication of the memory cost is that more obligatory syntactic components would consume more resources, implying that a sentence with less obligatory syntactic components would be easier to process. The integration cost is derived from the necessity of absorbing new input words into the existing syntactic structure. According to Gibson (2000), there are two steps of integration, and the first one is structural integration that requires a syntactic category to be matched with a syntactic expectation in the already-built syntactic structure. The second step of integration, as suggested by Gibson, is to comprehend the structural attachment resulted from the previous step. For example, a listener would evaluate whether a resultant discourse is plausible. Consider the following two example sentences extracted from Gibson (1998):

(a) The reporter who the senator attacked admitted the error.
(b) The reporter who attacked the senator admitted the error.

According to Gibson (1998, 2000), despite that the two sentences above are identical in length and words used, sentence (a) would impose a higher memory and integrate cost than sentence (b) because the object-relative structure put the argument “who” and the head “attack” at distance due to the intervening referent “the senator”. To process the verb “attacked” in sentence (a), a listener must keep “the reporter who the senator” in memory while connecting the verb to the already built syntactic structure. By comparison, processing the same verb in sentence (b) would be relatively straightforward as less information is held in memory and the syntactic structure is easier to build. Other scholars reported similar findings (e.g., Clahsen and Felser, 2006a; Felser and Roberts, 2007). Felser and Roberts (2007) conducted a study focusing on the L2 listening of long distances caused by wh-sentences and reported that L2 users handle long syntactic distance differently from native speakers in the way that their representations of the sentences lack abstract structures which could be the result of insufficient available WM resources or inadequate native-like processing mechanisms.

2.6.2.5 Greater Cognitive Cost in L2 Listening

Studies have found that auditory information processing can be more cognitive challenging for L2 English users. At the word level, although bilingual listeners with high L2 language proficiency can recognise words during speech processing as efficient as native speakers (Duyck, 2016; Grosjean, 2018), they seem to be more phonologically vulnerable than native speakers (Broersma and Cutler, 2008). Broersma and Cutler suggested that L2 listening can activate phantom words that are non-exist in the language and that the concurrent activation of candidate words induced by the phoneme sequence lasts longer and involves more words for L2 listeners than native speakers, and while native speakers can swiftly reject the near-words, L2 listeners would struggle with the spurious phantom lexical candidates that result in slow processing. L2 users are also more vulnerable to WFE than native speakers, not because of different processing dynamics but due to different levels of language exposure (Cop *et al.*,2015; Brysbaert *et al.*, 2017; Field, 2019). Kuperman and Van Dyke (2013) designated vocabulary size as the proxy of language exposure, with more words in the lexicon suggesting more significant exposure to a language, implying that interpreters working from their second language could face a steeper WFE.

At the sentence level, L2 parsing is less effective than L1 (Clahsen and Felser, 2006b; Jiang, 2004, 2007; Hopp, 2015). Sentence parsing involves two modes, in-depth parsing and shallow parsing (Sanford and Sturt, 2002; Clahsen and Felser, 2006a). According to Clahsen and Felser (2006a), in-depth parsing entails detailed syntactic representations as it receives substantial support from grammatical knowledge, whereas shallow parsing is mainly constrained by surface cues such as lexical-semantic information, and therefore, less grammatically demanding. Clahsen and Felser (2006b) reported that L2 users typically adopt shallow parsing more extensively than in-depth parsing. Specifically, L2 users are less effective in utilising syntactic cues and would compensate the deficit by resorting more to lexical-semantic information during sentence processing.

Similarly, Jiang (2004, 2007) reported that L2 users, such as Chinese students who learn English, show insensitivity to morphosyntactic formality errors. Based on the observation, a shallow structure hypothesis (SSH) claims that L2 users would construct representations that lack in-depth hierarchical structural analysis, and that native-like processing is limited to local sentence constituents (Clahsen and Felser, 2006a). The authors argued that shallow processing predominates L2 parsing regardless of their proficiency and that such inadequacy results from inadequate grammar knowledge and the lack of parsing mechanisms.

L2 users are also found to differ from native speakers in predictive processing. Mitsugi and MacWhinney (2016) proposed that inadequate L2 proficiency may impede predictive processing. However, Hopp (2015) suggested that L2 users only rely on lexical-semantic information instead of integrating it with morphosyntactic cues like native speakers do. Similar results are also reported by Kaan (2014). According to Hopp, L2 users are equipped with the knowledge to recognise morphosyntactic information, such as case marking, but they are limited to only utilising lexical-semantic cues because L2 processing is less automatic and more cognitively demanding. Another possible reason proposed by Hopp relates to the previously mentioned prediction-by-production mechanism. As Pickering and Gambi (2018) suggest, a listener anticipates the upcoming speech by imitating what has been heard so far and process it incrementally before running the representation in her or his language production system. Hopp believes that imitating and processing the temporal utterances make it more cognitively gruelling and susceptible to errors, causing L2 users to run the representation.

Another difference between L2 users and native speakers in sentence processing is pragmatics processing. In contrast to the first language, utilising world knowledge in processing a second language imposes a heavier burden on LTM as it requires more in-depth lexical research (Romero-Rivas *et al.*, 2017). It is previously reviewed that semantic memory, as a part of explicit memory in LTM, accommodates world knowledge and vocabulary and that retrieving information from semantic memory requires explicit efforts. Romero *et al.* conducted an ERP study where native and L2 Spanish speakers are presented with spoken sentences, some incongruent with word knowledge. The results showed that the violation of world knowledge triggers greater N400 negativity during L2 speech comprehension than L1 processing, suggesting that L2 comprehension demands deeper lexical search in contrast to L1.

Finally, it is suggested that L2 users are less effective as syntactic revision (Pozzan and Trueswell, 2015). As mentioned earlier, miscomprehensions arising from ambiguous utterances require rectification via re-analysis. Pozzan and Trueswell conducted an eye-tracking study to investigate how adult L2 learners process auditory garden-path sentences whose correct interpretation is usually the result of successful revisions of the initial incorrect interpretations. The authors reported that both native speakers and L2 users utilise referential information in the early processing stage and, therefore, are prone to show a garden-path effect. However, the two groups differ during revision. In contrast to native speakers, L2 users tend to be more profoundly challenged in recovering from misinterpretations. Specifically, compared to native speakers, L2 users show more difficulties in abandoning incorrect initial interpretations and are more error-prone in processing sentences with ambiguity. Pozzan and Trueswell attribute the L2 users' more substantial commitment to miscomprehensions induced by ambiguity due to insufficient available cognitive resources as L2 processing is more cognitively costly than L1.

2.6.3 Notetaking in CI

2.6.3.1 Overview of Notetaking

Peverly and Wolf (2019, p. 320) define notetaking as "the act of selecting and cryptically and idiosyncratically transcribing important information that can be used as a personal memory aid for later reference, review, and/or memorisation by the note taker". According to the authors, notetaking carries three features. Firstly, notes can store information from a particular activity (e.g., a lecture or a book) or reflect personal experiences. Second, notetaking in some scenarios, such as taking notes from a lecture, can be cognitively demanding. Third, notes are taken to fulfil different tasks. Traditionally, studies on notetaking have primarily focused on its two functions in learning, storage function and encoding function (e.g., DiVesta and Gray, 1972; Einstein, Morris, and Smith, 1985; Kiewra, 1985; 1987, 1989; Kiewra, Benton, and Lewis, 1987; Kiewra and Benton, 1988; Kiewra and Frank, 1988; Kiera, Mayer, Christensen, Kim, and Risch, 1991; Kiewra, Benton, Kim, Risch, and Christensen, 1995; Peverly and Wolf, 2019). However, the results yielded from these studies should be able to contribute to understanding the prominent role of notetaking in CI, as consecutive interpreters also take notes to alleviate memory pressure and to grasp the sense of the source speech (Sleskovitch, 1975, 1978; Albl-Mikasa, 2008, 2017; Pöchhacker, 2015, 2016; Chen, 2016; Gillies, 2017, 2019).

DiVesta and Gray (1972) conducted a seminal study on the role of notetaking in learning. In the study, the authors sat 120 subjects in an experiment where notetaking and review of the notes were manipulated. Performance on recall and multiple-choice tests were compared across various conditions. The results indicate that notetaking, compared with the condition where notetaking was prohibited, can reliably contribute to better recall and higher test scores. Besides, according to DiVesta and Gary, the facilitatory role of notetaking is particularly salient when note-reviewing is possible, as indicated by the results that the rehearsal of the presented information also enhances recall and multiple-choice test scores. The authors concluded that notetaking and note-reviewing are mathemagenic activities that boost learning. The authors attribute the facilitatory role of notetaking to its two functions, namely encoding function and external storage function (Kiewra in 1985 referred the encoding function as “process function” and the external storage function to “product function”) and suggest that notetaking facilitates learning by engendering either one or both of the two functions. The encoding function is hypothesised to enable an individual to process the verbal presentation during which the notetaker “transcribe(s) whatever subjective associations, inferences, and interpretations occurred to him while listening” (p. 8). The storage function is believed to serve as an external storage with which information can be reviewed.

DiVesta and Gary’s (1972) study paved the avenue for studies focusing on the role of notetaking in academic achievements. A dominant volume of such studies was published during the 1980s and 1990s, most conducted by Kenneth Kiewra and his colleagues. Apart from endorsing DiVesta and Gary’s proposal for the two functions, scholars who investigate notetaking in classroom settings tend to agree that although both functions are beneficial to learning, the storage function is superior to the encoding function in enhancing academic performance (e.g., Kiewra, 1985; Kobayashi, 2006; Armbruster, 2008).

In 1985, Kiewra extensively reviewed 56 studies investigating the encoding function of notetaking and found that most of the studies support notetaking as a facilitatory technique to learning. For example, it is found that compared with neglected information, noted information enjoys a better chance to be immediately recalled (Fisher and Harris, 1973; Bretzing and Kulhavy, 1981; Kiewra and Fletcher, 1984). Similar results were observed from studies where performances on delayed recall were compared. Several studies found that, following a period of delay, the likelihood of noted information to be recalled is dramatically higher than un-noted information (Howe, 1970a [34% versus less than 5%]; Aiken, Thomas, and Shennum, 1975 [47% versus 12%]). According to Kiewra’s interpretation of the results, while notetaking does not guarantee recall, the absence of notes could almost result in an inability to recall. However, a minority of reviewed studies failed to support the encoding function. For example, Kobayashi (2006) suggested that the value of notetaking cannot be effectively judged by only looking at the encoding function because benefits from notetaking, in academic settings, derive from viewing of notes. Similarly, Carter and Van Martre (1975) suggested that without the opportunity to review the notes, taking notes from a lecture bears no difference from mentally processing the speech in enhancing recall.

In the same study, Kiewra also explored the storage function of notetaking by reviewing 22 studies where note-review was manipulated. In the abovementioned study conducted by Carter and Van Matre, participants were instructed to either take notes or rely solely on memory when listening to a verbal presentation. To certify the efficacy of note-review in enhancing immediate and delayed recall, the notetaking participants were further divided into two groups where note-review was either allowed or prohibited before performing a recall task. Results indicate that participants who took and reviewed notes recalled the most information, whereas those who only listened to the lecture recollected the least information. Based on the data, Carter and Van Martre concluded that the benefits of notetaking are engendered by subsequently reviewing the notes instead of the behaviour of taking notes. Einstein, Morris, and Smith (1985) reported that although note-review yields little effect on boosting immediate recall, it significantly enhances delayed recall – participants who reviewed notes before the task recalled three times more information than their counterparts. Similar results were reported by other scholars such as Howe (1970b) and Pauk (1974). As suggested by the authors, an interesting aspect of note-review is that notes should be reviewed in close proximity to a specific task to ensure the best enhancement for recall. Based on the studies reviewed by Kiewra, while scholars voiced different opinions towards the benefits of the encoding function, they agreed that the primary benefits of notetaking reside in the storage function.

In 1991, Kiewra and his colleagues contributed to a new classification of notetaking functions (Kiewra, DuBois, Christian, McShane, Meyerhoffer, and Roskelley, 1991). Kiewra *et al.* believe one of the flaws in DiVesta and Gary's classification (1972) is that the storage function combines both encoding and storage. Specifically, an individual would first have to take notes (encoding) before reviewing them (storage). To independently compare encoding and storage, it is necessary to differentiate the two. Therefore, the authors suggested that notes offer three functions: encoding (notes are taken but not reviewed), encoding plus storage (notes are taken and reviewed), and external storage (absent from lecture but review borrowed notes). Participants were divided into three groups in the study based on the new classification and exposed to videotaped lectures. A cued recall test was performed to compare the efficacy of different notetaking functions. Results indicate that encoding plus storage surpasses the other two functions in facilitating recall, a notion which later studies lent support to (e.g., Armbruster, 2008; Thorly, 2016).

Some studies have shown efforts in explaining why the storage function is more facilitatory in recall and comprehension than the encoding function (e.g., Carter and Van Matre, 1975; Kiewra *et al.*, 1991; Kiewra, 2016; Poverly and Wolf, 2019). From the perspective of the encoding function, studies have revealed that the function is unreliable as notes usually omit a significant volume of information from the speech. Hartley and Marshall (1974) reported that notetakers, on average, record only 11% of presented information. Locke (1977) indicated a higher ratio (52%) but highlighted the dramatic 48% omission rate. O'Donnell and Dansereau (1993) suggested that notetakers recorded just 25% of the total number of ideas. More recently, Kiewra (2016) reported that across the 16 studies conducted by Kiewra and his colleagues investigating the completeness of notes, subjects achieved a range of completeness rate between 20 and 70%, yielding an average of 35%. Kiewra (2016) also indicated that notes contain inaccurate or vague statements apart from high omission rates.

According to Poverty and Wolf, several factors contribute to incomplete and inaccurate notes. To begin with, the dramatically incompatible speech rate and handwriting speed mean that notetakers cannot record or comprehend a significant amount of information in a meaningful manner. Studies have revealed that while speech rate could range from 120 to 190 wpm (Marslen-Wilson, 1973; Riding and Vincent, 1980; Rivers 1981), handwriting speed is less than 40 wpm (Greene, 1928; Piolat, 1982; Foulin, 1995; Summer and Catarro, 2003). Second, notetakers usually lack sufficient background knowledge of the topic. As a result, the recorded information is likely inadequate to make meaningful sense of the material. Third, verbal presentations usually contain high propositional density, increasing the difficulty in understanding the speech. Studies have found propositional density is a predictor of text readability – texts with higher propositional density are more challenging to understand (Kintsch and Vipond, 1979; Kintsch and Kintsch, 1998).

From the perspective of the storage function, the literature has explained why note-review generates primary benefits. According to Carter and Van Matre (1975), note-reviewing benefits recall from three aspects. First, notetakers can repetitively rehearse the recorded information by reviewing notes, contributing to consolidation. Second, reviewing notes helps concentrate on retrieval endeavours, sifting germane information from the noise. Finally, reviewing notes engenders a cueing effect that assists the recollection of noted and neglected information, compensating for high omission rates. However, the abovementioned benefits do not erase another drawback of the encoding function, which is the inclusion of inaccuracies. Peverly and Wolf (2019) suggested that reviewing notes may rectify the flaw. According to the authors, reviewing notes allows notetakers to add context and perform in-depth processing during which presented information is recalled and conceptualised.

Despite that most of the efforts in investigating notetaking are invested in classroom settings, the technique of notetaking is applied in non-academic environments and is found to inherit similar benefits (enhancing recall) and drawbacks (omission and inaccuracies), such as in trials (Kiewra, 2016; Thorley, 2016). As mentioned earlier, as a distinctive feature of CI, interpreters commonly adopt notetaking (Pöchhacker, 2014; Chen, 2016; Dawrant and Setton, 2016). CI notes are taken as a mnemonic device that facilitates recalling what has been understood from the speech (Seleskovitch, 1968; Albl-Mikasa, 2016). Although CI notes are not taken to enhance academic performances, scholars believe they are significant to output quality (Groot, 1997; Gile, 2009; Gillies, 2019). Despite the dearth of research, the past decades have witnessed sustained interests in understanding CI notetaking regarding its principles (e.g., Rozan, 2002; Gillies, 2019) and process (e.g., Ito, 2017; Someya, 2017; Albl-Mikasa, 2008; 2017). The rest of this chapter reviews CI notetaking in terms of principles, linguistic aspects, and cognitive aspects.

2.6.3.2 Systems and Principles of CI Notetaking

Since the 1950s, professional interpreters or instructors, based on practical interpreting experiences, proposed approaches that are understood as traditional notetaking techniques. These systems regard notes were often perceived as language-independent and are primarily prescriptive. The earliest and perhaps the most famous CI notetaking system was proposed by Rozan in 1956. Rozan’s system is regarded as the classical approach and has created momentous impacts and influenced later professionals, such as Gile (1997), Matyssek (1989), and Gillies (2017, 2019).

Rozan’s seminal notetaking system consists of seven principles. The first principle instructs interpreters to note ideas instead of words. According to Rozan, interpreters must focus on main ideas and take notes in a clear and straightforward manner. A similar idea was raised by Matyssek (1989). Gillies (2017, 2019) echoes with the principle when suggesting that symbols in notes should represent concepts rather than the actual words or expressions uttered by the speaker. In the view of Alexieva (1994), grasping ideas instead of words is attainable by analysing the source speech in terms of the role each component in contributing to the architecture of meaning.

The second principle in Rozan’s model includes rules on abbreviation. Notetaking is usually performed under severe time pressure, suggesting the need to shorten and condense information (Piolat *et al*., 2017). According to Rozan, the crux of the second principle is to avoid writing down whole words unless they are short (4 to 5 letters). For longer words, only the first and last letters should be noted, with the latter raised as superscript. For example, “*production*” should be noted as “*pron*”. The rule of abbreviation is also seen in other publications. Gillies (2017, 2019) suggested that, when applicable, interpreters should take advantage of shared suffixes to make abbreviations even more straightforward. For example, “*-tion*” should be noted as a raised “*n*”. Following Gillies’s rule, “*production*” would be noted as “*prodn*”. Also, according to Gillies, to write less, interpreters can deliberately use a misspelling or phonetic spelling as CI notetaking is not bound by spelling rules (Seleskovitch and Lederer, 1989). For instance, “*through*” can be written down as “*thru”, and “late*” can be noted down as “*l8*” (Gillies, 2017, p. 181). According to Rozan, abbreviations should also be applied to indicate tense and register. It is worth noting that the abovementioned rules on abbreviation are not universal to all languages. Gillies (2017, 2019) suggested that the use of misspelling or phenetic spelling would be futile to languages that are written phonetically, such as Polish and German. By contrast, the strategy would prove helpful for English or French users because phonetical spelling in those languages is usually shorter than correct spelling. In terms of the Chinese language, since characters are composed with strokes instead of alphabets, some aspects of the rule of abbreviation might not be helpful (Liu, 2008), although the rule still yields benefits, especially when a speech is interpreted into English (Chen, 2016). Despite the flexible ways to abbreviate, they shall not result in ambiguous or inaccurate notes (Schweda Nicolson, 1990; Rozan, 2002; Gillies, 2017, 2019; Alexieva, 1994).

The third rule highlights the noting of links. According to Rozan (2002, p. 18), “an idea can be distorted completely if it is related to the previous idea is not clearly indicated”, and, therefore, interpreters shall not neglect links when taking notes. To capture links, Rozan suggested using one keyword or symbol for a group of similar links. For instance, “*tho*” can be used to note links that indicate adversative relations such as “*although*” and “*despite that*”. Gillies (2017, p. 66 - 67) adopted a similar “one-for-all” idea, with some aspects being more refined. For example, a distinction is made between “contradiction or limitation following an idea” and “contradiction or limitation preceding an idea”, with the former (e.g., but, however, on the other hand) noted down as “*B*” and the latter (e.g., although, while, notwithstanding) noted as “*THO*”. Regarding the rule of noting links, Gillies made new development by calling for discretion in noting implicit links and ignoring pseudo links, such as “*and*” and “*so*”, which are often used as fillers.

The fourth and fifth rules within Rozan’s system suggested the noting of negation and emphasis. To represent negation, interpreters can cross a word or symbol with a line. To emphasise, depending on the level of certainty, interpreters can underline the word or symbol once or twice. Later works, such as Matyssek (1990) and Gillies (2017, 2019), inherited the use of lines in indicating negation and significance. Gillies further proposed that emphasis can be indicated by an exclamation mark or writing the weighted information on the right side of the notepad.

The sixth and seventh principles emphasise verticality and shift, respectively. According to Rozan, the two principles are the backbones that underpin the entire system. The two principles specifically deal with the layout of the notes and have been accepted by other scholars (e.g., Ilg and Lambert, 1996; Kohn and Albl-Misaka, 2002; Jones, 2014; Gillies, 2017). The technique of verticality indicates that notes should be taken from top to bottom instead of from left to right on a notepad, whereas shifting is referred to as “writing notes in the place on a lower line where they would have appeared the text on the line above been repeated” (Rozan, 2002, p. 22). According to Kohn and Albl-Mikasa, to practice verticality, interpreters shall structure notes in a “vertical indented and terraced way”, leading to a logical presentation of semantic relations between different parts (p. 262) and “complete and immediate synthesis” when reading notes (Rozan, 2002, p. 20). Gillies (2017, 2019) and Jones (2014) further suggested that main ideas should be noted in a subject-verb-object (SVO) structure and that the three elements should be arranged diagonally from the top left to the bottom right. This SVO structure would show the basic structure and information of a sentence and allow interpreters to add details.

Although the vertical presentation and shifting of notes have been widely accepted since its proposal by Rozan, the rationale behind the practice only came to light in recent years. Jones suggested that arranging notes in a diagonal manner firstly forces interpreters to divide a sentence into its components, avoiding semantic confusions or syntactic inference. Also, instead of presenting information side by side, it allows clear visual indication of new ideas.

Finally, notes taken diagonally are concise and sense oriented. Gillies added that the left-to-right diagonal layout is preferred also because “eyes move from left to right” (p. 44). However, this reason is less convincing not only because eyes can move from all directions but also because in some languages, such as Arabic, words are written and read from right to left, suggesting that interpreters shall be adaptive (Jones, 2014). Apart from indicating semantic relations, scholars on the subject of the layout of notes also advocate for separating ideas apart by using a horizontal line (Thiéry, 1981; Jones, 2014; Albl-Mikasa, 2017; Gillies, 2017, 2019). According to Jones, “a line should at least be drawn after each complete sentence” (p. 47).

Apart from prescriptively instructing interpreters on the content and layout of notes, existing notetaking systems also have proposed the use of symbols (Rozan, 2002; Matyssek, 1989; Gilles, 2017). According to Jones (2014), symbols enjoy several advantages that contribute to efficient notetaking. Firstly, symbols are timesaving to write and easy to read, meaning that interpreters would be able to process more information. Secondly, using symbols frees interpreters from source language interference as symbols represent ideas instead of words. Gillies (2017, 2019) added that symbols, being condensed in meaning and simple in form, can save space on the notepad and lead to tidy and structured note formats. Gillies also suggested that although the forms of symbols are not restrained to pictures or parts of words, the interpreter should ensure that symbols will represent underlying meanings or the concept of a word or phrase. For instance, verbs such as “suggest” and “propose” should share the same symbol because they are identical in concepts.

Once the rationale of using symbols is understood, it would seem reasonable to ask how much content in a set of notes should be symbols. Rozan (2002) proposed the modest or minimalist approach where 20 symbols are suggested, covering expression (e.g., thought), motion (e.g., transfer), correspondence (e.g., difference), and things (e.g., country). According to Rozan, among the 20 symbols, only 10 are indispensable. To explain such prudence, Rozan suggested that interpreters should not use too many symbols so that the notes will not be a collection of signs that need to be deciphered. Therefore, only frequently encountered concept words should be noted with symbols. The other end of the spectrum lies the maximalist approach represented by Matyssek (1989), who dedicated a whole book to listing detailed symbols and corresponding words. Ilg and Lambert (1996) criticised Matyssek’s approach for being too detailed, believing that novice interpreters tend to digest the approach as interpreter’s shorthand at the price of listening to the source speech. With a shared view with other scholars (e.g., Allioni, 1989; Gillies, 2017, 2019), Jones (2014) suggested that although it is a personal matter to find a position along the spectrum, interpreters should not create symbols in an arbitrary and complex manner. Instead, symbols must be unequivocal, immediately obvious, and organically systematic.

To sum up, as pointed out by Albl-Mikasa (2017), studies on the subject of systematic notetaking tend to agree on the following three basic principles: instantaneous seizability, and individuality. To cater to the principle of economy, interpreters ought to minimise effort by keeping notes “as scarce and brief as possible” (p. 72). The principle of instantaneous seizability highlights that notes must be instantly unequivocal. Otherwise, the pressure on interpreters’ WM memory system will not be alleviated. The principle of individuality, according to Albl-Mikasa, refers to the freedom of choosing systems or methods that best suit the individual interpreter’s need for memory support.

2.6.3.3 Linguistic Aspects of CI Notetaking

Although scholars and longstanding professional interpreters have reached a consensus on notetaking principles, their opinions on the linguistic aspects of notetaking have resulted in a controversy. Specifically, scholars’ opinions are divided on if notes should be a deverbalised entity (Seleskovitch, 1978; Seleskovitvh and Lederer, 1989; Matyssek, 1989) or if notes are linguistic (Kirchhoff, 1979; Ilg, 1980; Albl-Mikasa, 2017).

The notion that interpreters’ notes are language-independent is headed by Seleskovitch (1975; 1978). As previously mentioned in 2.6.2.2, according to Seleskovitch, successful interpreting entails a deep and comprehensive deverbalisation process (Seleskovitch, 1978 Seleskovitvh and Lederer, 1989; Setton, 2002; Lederer, 2010; Ito, 2017). According to the authors, the deverbalisation process requires interpreters to entirely and instantly dissociate sense from any language form in cognitive memory. As a result of the process, interpreters’ notes represent the ideas of the source speech and serve as memory triggers instead of “an exhaustive code” (Setton, 2002, p. 119).

The opposite pathway regarding the linguistic aspects of notes, pioneered by Kirchhoff (1979), concerns the linguistic dimension of interpreters’ notes. Ilg (1980) rejected Seleskovitch’s notion of instant dissociation from the source text and suggested that semantic features must be retained in some situations, such as formal international gatherings, where the formality of language matters. While endorsing the necessity for deep and comprehensive sense-seeking analyses, some scholars highlighted that interpreters' notes are not a set of deverbalised signs that are language-independent but rather an interpreter's individualised language (Kohn and Albl-Mikasa, 2002; Albl-Mikasa, 2017).

Based on a detailed comparison between natural language and notation language, Albl-Mikasa (2017) suggested that notation language can be seen as written language and that the difference between the two is "often only a matter of degree" (p. 80). According to Albl-Mikasa, the similarities between written and notational language sustain in three aspects. First, both written and notational language have consisted of symbolic signs. Natural language is believed to use sound, ideograms or letters, and symbols to discretely represent human knowledge (Ferrone and Zanzotto, 2020). It is suggested that expressions are produced using a wide array of linguistic signals, including sounds, words, phrases, and sentences (Kohn and Albl-Mikasa, 2002). Similarly, CI notes take advantage of a wide array of symbolic signs, such as whole words abbreviations, to create written (Albl-Mikasa, 2017) or verbal expressions (Kohn and Albl-Mikasa, 2002). Secondly, according to Albl-Mikasa, both natural language and notes are pictographic and ideographic in nature. Written languages are the results of cultural development that started with using cave paintings (signs) to express concepts before the logographic and alphabetic signs are assigned to sounds (spoken words). By comparison, interpreters' notes, due to the widespread use of symbols, enjoy the same nature but to a higher degree. Finally, Albl-Mikasa proposed that both written languages and CI notes are underpinned by mixed systems. Due to the development of written language, many natural languages use separate systems for words (also characters) and sounds. For example, some Japanese words can be written in *kanji* (logographic character borrowed from traditional Chinese) or kana (syllabaries that represent Japanese phonological units) (Rickheit and Strohner, 1993). In European languages, such as English, Arabic numbers and mathematical signs are often used to replace the spelt-out words. In comparison, CI notes are often known for their mixtures of linguistic and graphic features (Ilg and Lambert, 1996).

Apart from being highly similar to natural language in terms of composition, CI notes show linguistic features at the level of words, sentences, and discourse (Kirchhoff, 1979; Kohn and Albbl-Mikasa, 2002; Chen, 2016; Albl-Mikasa, 2017, 2019). At the word level, the linguistic feature of notes is firstly seen in the ability to express thoughts. It is believed that notation signs have lexical aspects that mimic natural language lexemes in a way that users can activate conceptual structures. Put simply, notation signs represent words (Kohn and Albl-

Mikasa, 2002; Albl-Mikasa, 2017). An apparent example would be the cases where a symbol is used to represent a word. The linguistic nature of notes is also embodied in the way notation language is formed and inflected. As one of the most commonly used means to achieve efficient notetaking, abbreviation takes advantage of the spelling of natural language words. It is suggested that notation signs can also be bounded using derivational or inflectional morphemes (Kohn and Albl-Mikasa, 2002). The use of derivational morphemes, or suffixation, allows interpreters to be highly creative in creating symbols. The use of inflectional morphemes helps interpreters to produce notes that accurately reflect gender, number, tense, mode, and case.

At the sentence level, semantic relations are visualised via the highly recommended vertical layout (see 2.6.2.2). Coordination of components, such as clauses, can be stacked together. Indentation should be given to subordinate clauses. Because the semantic structure of the source text is visually presented during notetaking, the authors suggested that the vertical layout of notes must be regarded as a notational way of expression.

At the discourse level, syntactic rules that dominate relations among components lose their significance, whereas the analysis of the speech structure or the construction of cohesion and coherence becomes crucial (Jones, 2014; Albl-Mikasa, 2017). Although cohesion and coherence of utterances primarily result from cognitive processing (Van Dijk and Kintsch, 1983; Sperber and Wilson, 1986; Wilson and Sperber, 2002; Albl-Mikasa, 2008; Johnstone, 2017), notes, despite being text-based, offer instant visual aids to interpreters in grasping links that connect different parts of the source speech (Jones, 2014; Albl-Mikasa, 2017; Gillies, 2017, 2019). To be specific, these authors suggested interpreters make use of the left-hand margin on the notepad as a space dedicated to noting cohesive markers (e.g., “because”, “however”) and global coherence indicators (e.g., “in summary”). Apart from noting cohesion and coherence, it is suggested that interpreters should use the left-hand margin to note critical points from the speech (Jones, 2014). Albl-Mikasa proposed that the margin is ideal for placing “attention-catching prompts” (Albl-Mikasa, 2017, p. 88) to remind interpreters of taking extra efforts in tackling specific issues during interpreting, such as stressing a nuance from the source speech. It should be noted that making use of the left-hand margin is not an arbitrary suggestion but rather a logical move because the left-to-right diagonal layout makes it convenient to indicate the link or reminder before each new idea (Jones, 2014).

To be succinct, linguistically speaking, CI notes contain wide-ranging lexical components, demonstrate derivational and flexional rules, indicate syntactic structures and semantic relations, and visualise speech structures (Albl-Mikasa, 2017).

2.6.3.4 Cognitive Aspects of CI Notetaking

Previous discussions have endeavoured to show that the listening phase of CI involves active listening and notetaking. The two activities entail deep and comprehensive mental analysis of the source speech to construct sense, the key to faithful rendition. To paint the whole picture and to fully understand the intricacy of active listening and notetaking, the cognitive aspects of the two activities must be explored. As previously reviewed, active listening and notetaking take place concurrently, and the latter records the result of the former in written forms. Given the inseparability, it would be reasonable to suggest that the cognitive aspects arising from listening comprehension are seamlessly transferred to notetaking, which would explain why scholars investigating cognitive aspects of CI notetaking (e.g., Seleskovitch, 1978; Albl-Mikasa, 2008; Setton, 2003; Someya, 2017) have typically focused on areas such as coherence building, mental representation, and sense construction during listening to the source speech (see 2.6.3.5). However, it shall by no means imply a complete overlap of the cognitive facets between active listening and the action of taking notes, as the latter has shown unique cognitive traits. To avoid repetition, this section is dedicated to reviewing the cognitive aspects exclusive to taking notes.

Perhaps the most distinctive cognitive aspect of CI notetaking is that it is a part of a cognitive process, and it consumes cognitive resources. While reaching a consensus on the stage where active listening and notetaking happen concurrently, scholars also agree that the two activities compete for the limited cognitive resource. The competitive relation among elements during consecutive interpreting is conceptualised in Effort Models proposed by Gile (1995, 2009). The models are built based on the assumption that interpreting requires cognitive resources in limited supply and that interpreting tends almost to deplete the resource pool and sometimes leaves it in a deficit that causes performance deterioration. Based on the Effort Model for CI, interpreting is conceptualised to contain two consecutive phases: the listening and notetaking phase; and the reformulation phase where target speech is produced. In phase one, on which this thesis is focused, interpreting is composed of four Efforts, namely Listening and Analysis, Notetaking, Short-term Memory operations, and Coordination. In phase two, efforts are invested in Remembering, Notereading, and Production. According to Gile, interpreters' processing capacity is only critical to the first phase. Effort Models are proposed for didactic rather than research purposes – they are set out to help professional and trainee interpreters to understand and overcome recurrent difficulties. The key message from the Models is that for interpreting to run smoothly, the required processing capacity must not exceed the available capacity (i.e., to avoid saturation) and that the processing capacity for a given Effort should not be insufficient (i.e., to avoid individual deficit). Gile suggested that notetaking constitutes a significant capacity-exhausting Effort, and over-investment in notetaking is usually responsible for performance deterioration. Therefore, interpreters should endeavour to "reduce processing capacity and time requirements of notetaking while maintaining the efficiency of notes as memory reinforcers" (Gile, 2009, p. 170). More specifically, Gile pointed out that cognitive and time resources spent on notetaking should be controlled at a minimal level to prioritise efficient Listening and Analysis.

2.6.3.5 Cognitive Efforts in CI notetaking

So far, this review of the literature on the listening and analysis phase of CI has shown that active listening requires a significant volume of cognitive efforts. Notetaking takes place in parallel with active listening and is considered as a significant cognitive recourse consumer during CI (Gile, 2009). However, the research on CI notetaking as a cognitive process has primarily focused on the processing of source speech, leaving a dearth in exploring the cognitive costs imposed by the action of creating CI notes. From the studies that investigate the cognitive prices imposed by notetaking regardless of its contexts and purpose, it is proposed that notetaking can impose a significant demand on WM resources (Piolat *et al*., 2005; Jansen, Lakens, IJsselsteijn, 2017).

Some scholars have shed light on the role that memory ability plays during notetaking from verbal presentations. DiVesta and Gray (1973) suggested that memory span affect the profitability of taking notes – notetakers with higher memory span are more likely to benefit from notetaking than their counterparts. Peverly (2006) proposed a similar finding that notetakers with higher handwriting rates tend to have a better memory of the speech content because faster notetaking allows a notetaker to record more propositions and listen to the speech more attentively. Kiewra and his colleagues correlated notetaking with the ability to hold and manipulate verbal information and suggested that the discrepancy in "WM skills" among notetakers could be manifested in the quantity and quality of notes. Specifically, notetakers with less competent WM skills, compared with their counterparts, recorded fewer words, total ideas (Kiewra *et al*., 1987; Kiewra and Benton, 1988), particularly subordinate ideas (Kiewra *et al*., 1987).

In recent years, developments (e.g., Piolat *et al*., 2005; Piolat *et al*., 2008; Jansen *et al*., 2017) have been made in investigating the cognitive costs imposed by notetaking. In a series of dual-task studies, scholars have compared cognitive activities in terms of their respective IRT (interference in reaction time), with longer IRT indicating higher demands imposed on the CE by the primary task. The tasks compared ranged from easy tasks such as copying texts (Olive and Piolat, 2002) to complex tasks such as planning (Piolat, Roussey, Olive, and Farioli, 1996). According to Piolet (2007), the IRT for notetaking exceeded 350 milliseconds, making it one of the most cognitively demanding tasks, surpassing playing chess (Britton and Tesser, 1982) and chasing text composing, translating, revising, and planning (Piolat *et al*., 1996).

Piolat *et al*. (2005) tapped into the cognitive demand imposed on WM system during notetaking and suggested that although schemata are involved in notetaking, such as accessing mental lexicon and forming words or characters, notetaking entails continuous metacognitive control of activities to keep grasped information in check, which adds difficulty to notetaking (Barbier, Faraco, Piolat, and Branca, 2004; Roussey and Piolat, 2005). Piolat *et al*. (2005) also indicated that notetakers are constantly faced with the time urgency induced by the vast difference between the speed of speaking and writing, which poses a particular difficulty in notetaking. To temporally manage information, the notetaker must maintain an active representation of the heard information in his or her WM system so that a portion of the information can be noted down and concurrently face constantly updated new information. Efforts are also required to coordinate multiple cognitive processing (e.g., listening comprehension), which adds pressure to the CE. According to Piolat *et al*., the time urgency of selecting and noting information while simultaneously handling new information imposes a significant load on the entire WM system, especially the CE. To circumvent WM overload, notetakers may face the choice to reduce the cognitive effort in either listening comprehension or taking notes.

From a more fine-grained perspective, notetaking as a process involves a recursive cycle composed of several activities. According to Jansen *et al*. (2017), to take note, the first step for the notetaker is to comprehend the material (listening comprehension), followed by identifying essential information, putting the information in context, paraphrasing or summarising, and finally recording information in written forms. It is worth reinitiating that in

CI notetaking, notes are usually in the form of partial words or symbols instead of complete phrases or sentences (Rozan, 2002). The five-step cycle of notetaking demonstrates that speech comprehension is one of the five components involved in the notetaking process. In other words, apart from understanding the speech, a notetaker must divide available attentional resources into four activities. Jansen *et al*. suggested that the cognitive load imposed by notetaking could lead to cognitive overload, which, given the individual differences in WM (Just and Carpenter, 1978; Daneman and Carpenter, 1980; King and Just, 1991), could partially explain the discrepancy regarding notetaking performance and benefits – notetaker with weaker WM abilities are more susceptible to cognitive overload (Jansen *et al*., 2017).

From the perspective of how cognitive load is constructed, notetaking firstly imposes intrinsic cognitive load as the result of handling the inherent difficulty of the speech, which is reviewed in 2.2.2. To recap, rare words consume more cognitive resources to recognise and subsequently retrieve their meanings from the mental lexicon; Complex syntax requires more mental efforts in parsing the sentence structure; and poorly structured speech demand more resources in extracting main ideas and comprehension. Extraneous cognitive load during notetaking is caused by several factors (Jansen *et al*. 2017). As mentioned in 2.2.2.1, activities that cause split attention during a primary task induce extraneous cognitive load (Chandler and Sweller, 1991, 1992; Mayor and Logan, 2014; Schroeder and Cenkci, 2018). Pertaining to CI notetaking, a split attention effect could be triggered when an interpreter needs to take notes from the speaker while visually processing slides. Jansen *et al*. (2017) suggested that cognitive load could also be affected by different note structures. In specific, notes taken in an organised structure are more cognitively consuming than transcriptional notes. The critical difference between the two types of note structures is whether to copy what is heard or to link new information with previously noted content, and the latter is believed to be more cognitively demanding than the former. As discussed previously, interpreters shall adopt vertical and diagonal layouts and indicate relations on semantics and discourse level.

2.6.3.6 Greater Cognitive Cost in L2 CI Notetaking

The nature of interpreting requires interpreters to work between languages. As previously reviewed in 2.6.2.5, L2 speech processing, compared with one's native language, is more cognitively demanding due to less automatised processing skills (Piolat *et al*. 2008), and such disadvantage is believed to sustain even in advanced L2 users(Clahsen and Felser, 2006a).

Since notetaking profoundly relies on successful listening comprehension, taking notes from a foreign language would arguably be more challenging than from one's mother tongue (Faraco, Barbier, and Piolat, 2002; Barbier and Piolat, 2005; Piolat, Barbier, and Roussey, 2008). Piolat *et al*. (2008) suggested that the lack of automatised processing skills leads to extra consumption of resources, thereby significantly hindering the metacognitive control of activities during notetaking. According to Faraco *et al*. (2002), taking notes for an L2 speech is more time-consuming. On the word level, the less automatic decoding process in L2 notetaking leads to longer pauses when detecting intra- and inter-words boundaries. In terms of speech content, Faraco *et al*. suggested that L2 users require twice as long to anticipate the organisation of the upcoming information. Also, it is suggested that compared with L1 notetaking, sifting key information while holding the syntactic hierarchy in mind in L2 notetaking is much more complex and susceptible to significant loss of information (Chaudron, Loschky, and Cook, 1994; Clerehan, 1995). Faraco *et al*. (2002) highlighted that taking notes from L2 adds additional cognitive load to the task

In short, CI notetaking can put a hefty pressure on interpreters' WM system, not only because speech processing can be inherently challenging but also because interpreters must deal with extraneous cognitive load resulting from having to arrange notes in an organised layout and condense selected information into symbols or partial words. When Interpreters need to work from L2 into L1, the cognitive costs in notetaking can be amplified – L2 notetaking imposes an additional cognitive load, making it more time consuming in general and less efficient in selecting main ideas.

2.7 Summary of Key Literature

Human beings rely on the cognition system to deal with all forms of cognitive activities (Sweller *et al*., 2011). It is generally accepted that the human memory system lies at the core of the human cognition system, and Atkinson and Shriffin's three-store memory system (1968) is a popular way to conceptualise the memory system consisting of the SR, STM, and LTM. SR is a minimum capacity that scans and picks up the stimuli. STM, albeit a larger capacity than SR, can only hold 5 to 9 digits. Without rehearsing, stimuli stored in STM would quickly fade. LTM stores information relatively permanently with its unlimited capacity. LTM consists of explicit or declarative memory and implicit or non-declarative memory (Graft and Schacter, 1985; Squire, 1992; Baddeley, 2009; Vakil *et al*., 2018; Wang, 2020). Explicit memory is where specific events and facts are stored, and it is explicit in a way that retrieving information needs conscious efforts, whereas implicit memory accommodates skills and procedures that one has learnt can apply subconsciously to perform a task (Cohen and Squire, 1980; Schacter, 1987; Baddeley, 2009). Explicit memory is further divided into semantic and episodic memory (Tulving, 1972), with the former storing world knowledge, such as linguistic knowledge and schemata and the latter storing specific events. By contrast, implicit memory is found to support priming (Baddeley, 2009; Race *et al*., 2019).

In 1974, Baddeley and Hitch proposed WM based on STM. Compared with STM, WM can hold and manipulate information of varying types towards complex cognitive activities. Similar to STM, WM is also a limited capacity. WM contains four components. The CE is the dominating component regulating the attention resources. The two slave systems are the VSSP and the PL, in charge of visual and audio information, respectively. The EB is added to the model relatively recently, and it is a buffer where information across WM stores and between WM and LTM is linked and bound (Baddeley, 2000). The amount of WM resources invested towards cognitive tasks is understood as cognitive load. The dominant perspective regarding cognitive load construction is that the total load is determined by the combined intrinsic, extraneous, germane cognitive load (Sweller *et al*., 1988; Sweller and Chandler, 1994; Sweller, 2011; Galy *et al*., 2012). The intrinsic cognitive load cannot be altered as it is imposed by the material. The extraneous cognitive load can be manipulated by the manner of presentation and the extra activities one must perform alongside the main task. Germane cognitive load is facilitatory as it helps subjects apply schemata or strategy, but for the germane cognitive load to happen, there must be enough available WM resources.

Measuring cognitive load can be realised via subjective, objective, or physiological methods. However, a visible indicator of a certain level of cognitive load is the behaviour of looking away or gaze aversion (GA) (Glenburg *et al*., 1998; Doherty-Sneddon *et al*., 2002). GA is hypothesised to be an overt strategy in response to cognitive difficulty and to enhance performance by allowing subjects to disengage from the distractions (such as human faces) so that more attentional resources can be concentrated on the cognitive task at hand.

Although eye movements (saccade and fixations) have been found to indicate high cognitive costs during interpreting, GA has not yet been explored, especially in CI active listening and notetaking. To achieve faithful rendition, interpreters must perform active listening (Gile, 2009; Setton and Dawrant, 2016). Although comprehending speech have top-down and bottom-up aspects, active listening is mainly attained through bottom-up processing. According to the blueprint for listeners (Cutler and Clifton, 1999), bottom-up processing involves four stages, namely Decoding, Segmentation, Recognition, and Integration, each of which requires cognitive efforts. A number of factors can enhance the difficulty of listening. Uncommon words (Preston, 1935; Brysbaert *et al*., 2018; Popov and Reder, 2020) and syntactic complexity (Just and Carpenter, 1992; Ferreira *et al.*, 1996; Gibson, 1998, 2000) have been proven to be capable of imposing high cognitive load. Since source speech segments usually overwhelm memory capacity, CI interpreters usually resort to notetaking. Studies on notetaking suggested that notetaking executes encoding and storage functions (DiVesta and Gray, 1972), with the latter being more facilitatory in recall and comprehension because it allows the reviewing of notes (Kiewra *et al.*, 1991; Kiewra, 2016; Poverly and Wolf, 2019). Over the years, several CI notetaking systems and principles have been published, among which Rozan's system (Rozan, 2002) is accepted as the classic approach. Despite its role in alleviating memory pressure, CI notetaking is a cognitively challenging task as it vies for WM resources with listening, requires decision making, and must be performed under straining time urgency (Piolat *et al*., 2005; Gile, 2009). Active listening and notetaking in CI can be more cognitively challenging if the speech is delivered in interpreters' second language, for that L2 processing is less automatic and more cognitively demanding (Hopp, 2005; Pozzan and Trueswell, 2015; Romero-Rivas *et al*., 2017).

Given the likelihood of cognitive overload during active listening and notetaking in CI and the function of GA as a high cognitive load indicator, it seems warranted for this project to explore if the GA paradigm applies to CI.

Chapter 3 Initial Corpus Study

This chapter presents the Initial Corpus Study. The general aim of this study is to construct and code a corpus to establish a fact that interpreters would conduct GA during CI active listening and notetaking. Research design is reported in 3.1 where an exhaustive procedure for constructing and coding the corpus can be found. Results are presented in 3.2 and discussed in 3.3. The chapter ends at 3.4 where a short conclusion of the study is presented.

3.1 Methodology

3.1.1 Aims and Research Questions

3.1.1.1 Research Aims

The overall purpose of this corpus study is to investigate GA as a cognitive load indicator in the listening and notetaking period in CI. As mentioned in 2.3.2, GA, the behaviour of looking away is found to be an overt strategy for managing high cognitive load induced by difficult questions. So far, GA is studied predominantly in conversational settings where a person would look away from the interlocutor's face when answering difficult questions. The reasons that underpin the GA during conversation come in two aspects. First, the human face contains rich information, and decoding such information consumes cognitive resources (Glenberg *et al.*, 1998; Bruce and Young, 2012) and distracts the person from the task at hand. Second, looking away from the interlocutor's face would help the person disengage from the distraction and channel more cognitive resources to processing the question. Interestingly, there is a similarity between face and CI interpreters' notes – they both hold abundant information. While the face signals various attitudes and emotions, CI notes hold a myriad of information that reflects the contents and sense of the source speech. This similarity raises the question if GA would find its ground in interpreting studies.

Scholars in interpreting studies agree that interpreting is cognitively challenging because interpreters must divide attentional resources into multiple sub-tasks. In consecutive interpreting, cognitive load is mainly constructed during active listening and notetaking (Gile, 2009). An interpreter would need to concentrate on fully comprehending the speech while simultaneously noting down information to record the results of active listening and, therefore, is exposed to high cognitive load. In recent years, research on interpreting from a psycholinguistic perspective began to use eye-tracking technology to investigate how interpreters' eye movements might indicate an increase in cognitive load (Seeber, 2012, 2013; Seeber and Kerzel, 2012; Stachowiak-Szymczak, 2019; Tiselius and Sneed, 2020; Lin, 2022).

For instance, Stachowiak-Szymczak (2019) used the eye-tracking technology in studying interpreters' cognitive load and suggested that eye movements correlate with language processing during interpreting. According to Stachowiak-Szymcza, fixation durations were the longest when interpreters process problem triggers while looking at congruent pictures, a finding that lends support to the suggestion that higher cognitive load leads to longer fixations (Carpenter and Just, 1980). However, published studies on conference interpreting have focused on fixations and saccade, leaving a paucity for GA to be explored.

To fill the gap in research and bring GA and interpreting studies together, this study attempts to explore interpreters' GA during CI active listening and notetaking. Specifically, the corpus study is carried out to fulfil two aims. The first is to establish that interpreters do avert their gaze while interpreting. As previously discussed, GA has not yet been studied in conference interpreting. It is necessary to provide evidence that interpreters look away from their notepads during active listening and notetaking to break the ground for the current study to be justifiably conducted. On the premise that GA is an expected behaviour for CI interpreters, this corpus study is keen to investigate if GA is accompanied by the action of taking notes or listening only.

The second aim of this corpus study is to explore potential linguistic factors that lead to GA during active listening and notetaking and if the linguistic factors would contribute to the increase in cognitive load experienced by interpreters. Previous research confirmed that GA happens at critical moments when cognitive load is high during conversations but yielded little information on the relationships between GA and linguistic factors in CI settings. Studies on conference interpreting (e.g., Gile, 2009; Korpal and Stachowiak-Szymczak, 2020) have reported problem triggers that increase processing difficulty, such as complex syntax, numeric, and speech rates. Pursuing the second aim would help establish whether GA could reflect interpreters' processing difficulties.

3.1.1.2 Research Questions and hypotheses

As mentioned in the previous section, although GA is reported to indicate high cognitive load during conversations, its ground in CI settings is yet to be broken. To pursue the first research aim - to establish evidence for GA in CI settings, the following research questions are proposed:

RQ1: Do interpreters look away from notepads during active listening and notetaking in CI?

RQ2: Will interpreters engage in notetaking or detach from notetaking when they look away?

Exploring potential linguistic factors that induce GA, which is the second aim of this corpus study, leads to the following research question:

RQ3: What linguistic elements from the source speech would increase interpreters' cognitive load and lead to GA?

Based on the relevant literature reviewed in Chapter 2, hypotheses are proposed for the above RQs. For RQ1, it is hypothesised that interpreters will look away from notepads during active listening and notetaking in CI due to high cognitive load. The hypothesis is backed by the fact that active listening and notetaking are resource-intensive (Cutler and Clifton, 1999; Gile, 2009). To reiterate, active listening in CI requires interpreters to give uncompromising attention to every detail of the source speech for faithful rendition and primarily relies on bottom-up processing that entails several subtasks sensitive to problem triggers. Processing the source speech is usually accompanied by notetaking, which helps interpreters alleviate pressure on the memory and record content and the sense of the speech (Seleskovitch, 1978; Albl-Mikasa, 2008; Setton, 2003; Pöchhacker, 2016; Someya, 2017). Although taking notes facilitates information storage and encoding, it consumes cognitive resources by propelling the interpreter to fulfil several tasks such as contextualising and paraphrasing information. Besides, the time urgency of selecting and noting information while simultaneously handling new information imposes a significant load on WM (Piolat *et al.*, 2005). Given the high cognitive cost of CI and GA's function as a cognitive load indicator, it would seem warranted to hypothesise that interpreters would look away from notepads.

For RQ2, it is hypothesised that there will be two scenarios regarding GA and its accompanying action: (1) interpreters will look away from their notepads while engaging in the action of handwriting; (2) interpreters will look away from their notepads without handwriting, i.e., listening only. The hypothesis is produced on two bases. First, interpreters are trained to take notes and are capable of taking notes without monitoring their handwriting - a finding from the author's MA Dissertation (Guo, 2016) where trainee interpreters from Newcastle University were instructed to take notes without monitoring their handwriting to see if shutting down visual feeds from notepads would result in a better understanding of the source speech. The subjects reported that taking notes without monitoring handwriting is manageable and resulted in fewer notes and a better understanding of the source speech. Second, as previously reviewed, cognitive load is not static - it fluctuates through the course of the activity, and such fluctuations are embodied in physiological measurements (Paas and van Merriënboer, 1994b; Doherty-Sneddon *et al.*, 2002; Paas *et al.*, 2003; Brünken *et al.*, 2010; Xie and Salvendy, 2000; Whelan, 2007). In the context of CI notetaking, given the necessity of the activity, the fluctuation of cognitive load could lead to GA with writing when disengaging from monitoring handwriting would suffice to spare enough cognitive resources for comprehension or GA without writing when interpreters must fully detach from taking notes to focus on active listening.

For RQ3, it is hypothesised that although the established problem triggers during interpreting would impact interpreters' GA, LFWs, syntactic complexity, or a mixture of both would be associated with more GA. According to Gile (2009), it is common for interpreters to feel baffled by a term or a sentence regardless of their experience. As reviewed in section 2.6.2.4, processing rare words and complex syntactic structures can overburden WM system. Compared with ordinary words, LFWs impose higher cognitive load because they are buried deeper in the mental lexicon and require more effort and time to process (Forster, 1976, 1981; Keuleers *et al.*, 2010; Brysbaert, *et al.*, 2018; Vitevitch *et al.*, 2018). In terms of syntactic structure, processing a source language that is syntactically very different from the target language, such as English and Chinese, would increase the consumption of resources because an interpreter would have to hold some syntactic information in memory while mentally computing it towards sense in the target language (Gile, 2009). Also, some syntactic structures, such as object-relative clauses, are notoriously pricey to WM (Just and Carpenter, 1992; Ferreira *et al.*, 1996; Gibson, 1998, 2000; Fedorenko *et al.*, 2004; Fallon *et al.*, 2006). When interpreters work from L2 to L1, the cost of dealing with rare words and complex syntactic structures is even higher because L2 processing is less efficient and straightforward (Clahsen and Felser, 2006b; Jiang, 2004, 2007; Hopp, 2015; Duyck, 2016; Grosjean, 2018). In the context of CI, dealing with LFWs and syntactically complex structures would require extra resources and result in higher cognitive load and more GA.

3.1.2 A Quantitative Approach

This corpus study adopted a quantitative approach. According to Stockemer, Stockemer, and Glaeser (2019), quantitative methods, by utilising descriptive and inferential statistics, helps researchers numerically describe a phenomenon and determine relationships between variables.

In this study, adopting a quantitative design most directly resulted from the quantifiable data, such as GA frequency (RQ1), proportion of GA type (RQ2), and proportion of textual features (RQ3). Stockemer suggested that quantitative research designs facilitate studies to test established theories with new data, by which the scope where the theory can be applied can be redefined, a notion that aptly summarises the overall purpose of this thesis, i.e., to explore GA in the sphere of interpreting studies with new types of data. Also, quantitative methods have been adopted in previous studies investigating GA (e.g., Doherty-Sneddon *et al.*, 2012). In specific, within the context of this corpus study, it would be warranted to suggest that whereas RQ1 would be adequately addressed by descriptive data, RQ2 and 3 contain aspects that require inferential analyses (e.g., whether there is statistical significance between different types of GA (RQ2) and if GA is significantly more sensitive to certain textual features (RQ3).

To cater to the design, data was generated and collected from three tasks: processing and coding the corpus, observing the textual features preceding and during GA, and scoring sentence difficulty. Among the three tasks, the corpus acted as the master task, under which the second and third task was conducted as follow-ups to ensure a more in-depth understanding of GA. Specifically, the descriptive data towards RQ1 and RQ2 was yielded from coding the corpus and subsequently analysing the mark-ups (See 3.1.2.4.2 for definitions). The data towards the inferential aspects of RQ2 exploited the GA measurements including GA counts and lengths. For RQ3, the descriptive data concerned the textual features preceding and during GA and, therefore, entailed identifying textual features that could potentially increase cognitive cost during CI active listening and notetaking. The inferential data for RQ3 was obtained through a scoring system (See 3.1.5.3 for the scoring system) for sentence difficulty.

Figure 2 below represents the overall process, data yielded, and the RQ that the task was designed to explore.

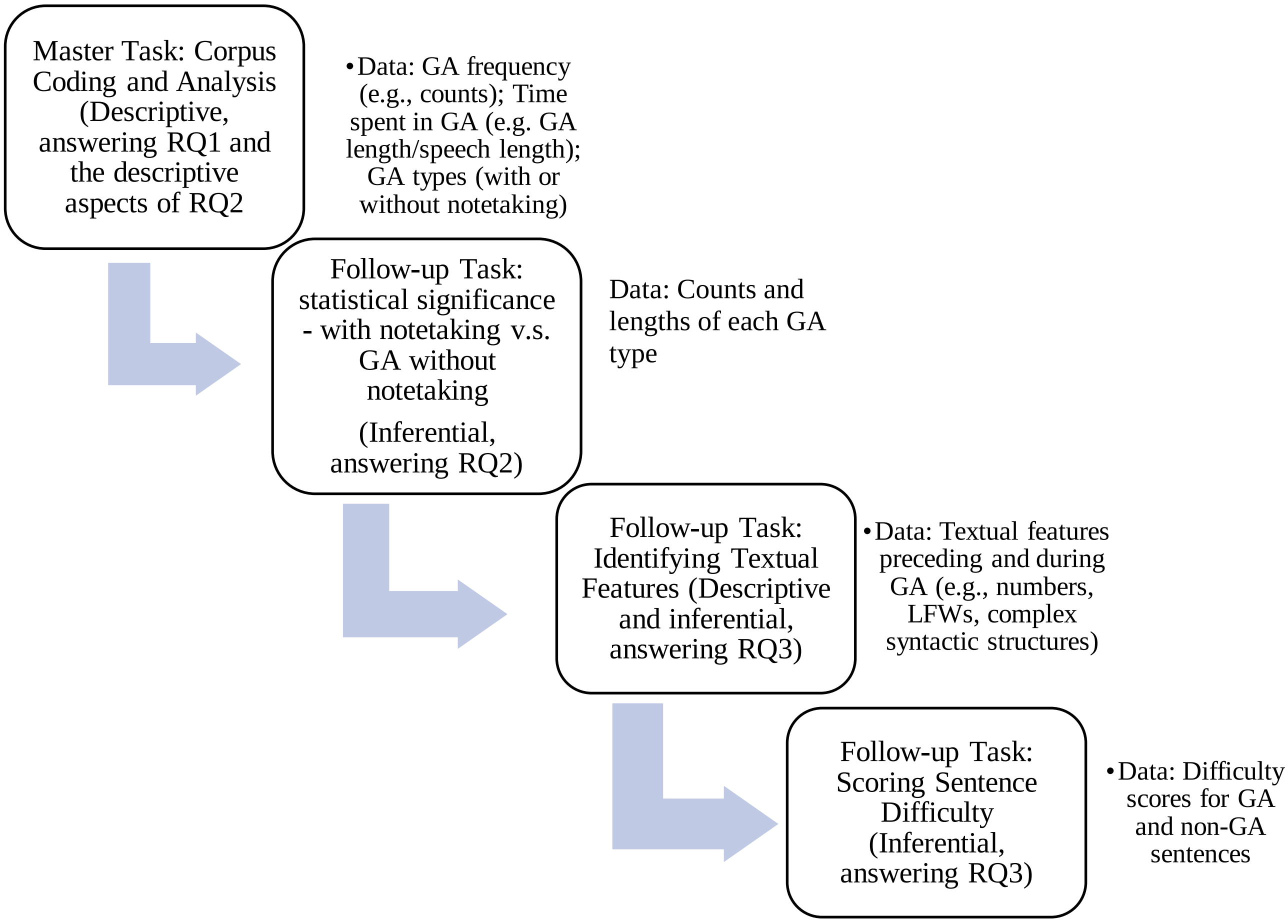


Figure 2 Data type and overall process of analysing data

3.1.3 Designing and Building the Corpus

Over the years, studies in corpus-linguistics have suggested critical considerations regarding the general principles (e.g., Atkins, Clear and Ostler, 1992; Lapadat and Lindsay, 1999; Wynne, Sinclair, and Leech, 2000; Sinclair, 2005; Adolphs and Knight, 2010) and specific aspects (e.g., Cheng, Creaves, and Warren, 2008; Reppen, 2010) during the designing and building of a corpus. According to Atkins *et al*. (1992, p. 2), corpus building involves five principle stages, namely "Specifications and design; Hardware and software; Data capture and mark-up; Corpus processing; Corpus growth and feedback". While the indications of most stages are evident, some would need clarification. "Specifications", according to Atkins *et al*., refers to establishing the type of corpus. For example, a corpus constructer would consider if he or she aspires to design and build a written or spoken corpus. For the current study, the specification of the corpus is spoken because the source materials are speech. "Corpus processing" refers to using tools to process the corpus and generate valuable data such as part-of-speech labels and word frequency. The following sections are dedicated to illustrating the process of building the corpus.

3.1.3.1 Recordings

Adolphs and Knight (2010) suggested that a fundamental stage of a spoken corpus is to prepare raw materials via recording. For the current corpus study, recordings were already available - video recordings of CI final examination (English to Mandarin Chinese) for stage 1 trainee interpreters (interchangeable with “interpreter” in this study) from the MA Programme in Translation and Interpreting at Newcastle University where the current study is carried out. Each video features a trainee interpreter performing active listening accompanied by notetaking and followed by target speech production. These videos contain rich information about the source speech and interpreters’ behaviours and, therefore, are well suited to the context of the current corpus study.

The current corpus study commenced in early 2018, and the available videos were recorded between 2013 and 2016. With over 40 Stage 1 students each year and two language directions (English to Chinese and vice versa) tested per exam, the video repository is enormous. Adolphs and Knight suggested that selecting raw material should follow practical criteria. For this corpus study, the selection of videos was based on the study's general aims and scale. To recap, the general aim of the study is to explore interpreters' GA during active listening and notetaking. The existing literature has suggested that GA is associated with high cognitive load, indicating that the videos should contain source speech with the potential to maintain cognitive load at a higher level. Considering the interpreters are all native Chinese speakers, potential videos were shortlisted from English to Chinese speech because processing L2 would impose higher cognitive load than L1 (see 2.6.2.5). However, the shortlisting does not suggest that the Chinese-to-English direction does not expose interpreters to high cognitive load.

Another significant factor to consider when selecting videos was whether the subjects' eye movements could be observed. For eye movements to be captured on camera, several conditions must be met simultaneously: (1) resolution must be good enough to capture eye movements; (2) the camera should be placed at an angle that allows a clear shot of the face of the interpreter; and (3) the eye area must not be obstructed by make-up or optical frames. The direct impact of meeting all three conditions was that some videos became unsuitable for the study. Although recording the process of CI final exams has been a common practice at the institute, it is not carried out to fulfil research purposes but as a means of documentation. Some videos did not meet the conditions and, therefore, were excluded.

As suggested by Adolph and Knight (2010), it is ethically significant to seek consent from people whose videos are to be coded in a corpus. The corpus study commenced in early 2018, meaning that the subjects in the videos had already graduated from their programme. A written consent form was sent to relevant graduates via email, and those who did not respond or agree to participate were not included as potential subjects. Based on the criteria and consent, 30 videos were selected, with 9 from 2013, 5 from 2014, 7 from 2015, and 9 from 2016. Interpreters from the same year interpreted the same speech, meaning that there are four source speech in total. Table 2 as follow is a summary of the features of all source speech.

Table 2 Features of source speech across four years

Year	***Speaker***	***Topic***	***Length (seconds)***	***Word Count***	***Speech Rate (wmp)***
2013	Speaker 1	Fiscal Cliff	222	528	143
2014	Speaker 1	Aqua Detox	218	458	126
2015	Speaker 2	Food Crises	194	502	155
2016	Speaker 1	Social Mobility	232	452	117

As indicated by the table above, the four source speech vary in length, speech rates, topics, and the speakers. Such differences are expected because the source speech is prepared for assessments instead of research, and it is within the nature of a corpus study to find patterns in materials with varying features (Stuart, Botella, and Ferri, 2016).

To give a clearer picture of the video content, it is necessary to briefly describe the process of the examination. The CI exams were conducted individually, meaning each video features only one interpreter. Interpreters were informed of the topic a week before the exam to gain familiarity with the topic. However, they were forbidden from bringing notes or any aid to the exam, and they were only allowed to bring a pen.

On the exam day, interpreters arrived at the venue at their allocated time slot. A reporter notepad was provided for interpreters. The notebook was shared among interpreters, but it was opened to a clean page for each interpreter. During the exam, interpreters were not allowed to go through the notes produced by previous interpreters. Once the interpreter was ready, the source speech was played through a speaker. Each source speech was divided into three segments, lasting between 1 and 1.5 minutes. After each segment, the interpreter was given some time to produce the target speech in Mandarin. All source speech was only played once.

3.1.2.2 Hardware and Software

The camera used to film the videos was a SONY Camcorder. The distance between the student and the camera was two meters. At the post-recording stage, the hardware used to build the corpus was a MacBook Pro laptop running a macOS X, version 10.14. The building of the corpus involved a variety of software installed on the laptop. The software can be divided into three groups: source speech editing, transcription, and coding. Table 3 lists the names and functions of the software used in this study.

Table 3 Software and functions

	Name of software	***Function in General***	***Functions in this study***
1	Audacity	Audio processing and editing	Exports source speech audios as WAV files
2	VLC	Multimedia player; audio/video editing platform	Strips audio from video
3	SPPAS	Adding annotation to audios and photos	Transcribes source speech
4	Praat	Performs spectrograms analysis and transcription	Adjusts word boundaries in source speech
5	ELAN	Facilitates annotating multimedia files	Allows adding mark-ups to the corpus
6	QuickTime	Streams multimedia files	Streams videos at half speed during corpus coding

Editing the source speech was achieved by utilising VLC and Audacity, a popular audio editing software in spoken corpus studies (e.g., Nelson, 2011; Supriya and Handore, 2017;

Ellis, 2019). In this corpus study, Audacity was used to crop the introduction at the beginning of the exam and the intervals between two adjacent speech segments as they were not a part of the source speech. VLC, another most used software in spoken corpus studies (Lai and Chen, 2015), was used to isolate the source speech from the video as the original speech recording for the 2013 exam was lost. According to Reppen (2010), naming conventions should be established before saving a file, and the name should reflect the file's content and be easy to group.

The edited source speech was subsequently transcribed, an indispensable step for a spoken corpus. It is generally accepted that transcribing speech into text format is time-consuming and laborious. Therefore, constructors should develop the awareness of using tools to make transcription more efficient and automated (Reppen, 2010; Adolphs and Knight, 2010). For this corpus study, transcribing source speech was relatively straightforward because the written format of the source speech was readily available. However, the written speech still needed to be digitised and tokenised. Transcribing the source speech was achieved via SPPAS, a user-friendly tool used to perform various annotations and segmentations on recorded speech (Bigi, 2012, 2014). This study used SPPAS for two functions: IPUs Segmentation and IPUscriber. The IPUs Segmentation function automatically segmented speech and silence, and IPUscriber allows each speech segment to be transcribed. The audio of each segment can be replayed using the control panel. The transcribed speech was tokenised to allow the corpus to be processed, such as adding part-of-speech (POS) tags and word frequencies. The software used for speech tokenising was ELAN and Praat. ELAN was mainly used as the platform where the corpus was compiled and coded. However, the software was also utilised to tokenise the transcribed source speech with its semi-automatic segmentation function – ELAN is equipped with a silence recogniser that facilitates the detection of the boundary between words (Durand, Gut, and Kristoffersen, 2014). However, the detection of word boundaries by ELAN can be inaccurate, and therefore, Praat, a popular tool used for building spoken corpus (Goldman, 2011; Boersma, 2014), was used to adjust the word boundaries manually. After adjustment, the tokenised source speech was imported back to ELAN, which marked the finishing of the corpus build-up.

ELAN, a multimedia annotation tool, is the platform where the corpus was compiled. It is developed to annotate audio and video files in psycholinguistics (Brugman, Russel, and Nijmegen, 2004). ELAN has been utilised in studies investigating gestures during linguistic activities (e.g., Wittenburg, Brugman, Russel, Klassmann, and Sloetjes, 2006; Lausberf and

Sloedjes, 2009; Pouw and Dixon, 2019), which has set a precedent for this corpus study. More importantly, the software was chosen for practical reasons. First, ELAN operates on a tier-based system, each tier being a container of annotations (Durand *et al*., 2014). The tier-based system generated excellent suitability for this corpus study as each interpreter's video needed to be coded individually. The tiers are arranged vertically so that visually quantifying mark-ups can be easily achieved. Second, ELAN has high temporal resolution – users can zoom in the recording spectrum to select the precise onset and finish moment of an action, with the maximum precision for each annotation to be one millisecond (Wittenburg *et al.*, 2006). Third, the software allows several annotation formats to co-exist, which is facilitatory to corpus construction as compiling a corpus usually involves a series of tools (Adolphs and Knight, 2010). Finally, ELAN has a search function that allows researchers to extract mark-ups as a tab-delimited file that can be processed further with other tools, such as excel (Wittenburg *et al*., 2006).

3.1.2.3 Step-by-step Procedure

To the author's best knowledge, the current study constitutes the first attempt in investigating GA via building-up and coding a corpus. To cater to the possibility of replication, a step-by-step procedure of building the corpus is provided in the following section. The following is a step-by-step process of building up the corpus using the videos from the 2013 exam:

- Step 1: Start by creating a new folder and name it SSEC2013 (Source Speech English to Chinese 2013).
- Step 2: Use VLC to convert audios into .wav format.
- Step 3: Import the .wav file from last step to Audacity where long intervals are trimmed to allow easy view of data in the corpus. Save the trimmed audio file as SSEC2013.wav and put it in the SSEC2013 folder.
- Step 4: Open SPPAS. From the interface, click "add file" then choose the source speech file SSEC2013.wav. The added source speech file will appear on the left side of the interface (See Figure 3 below).

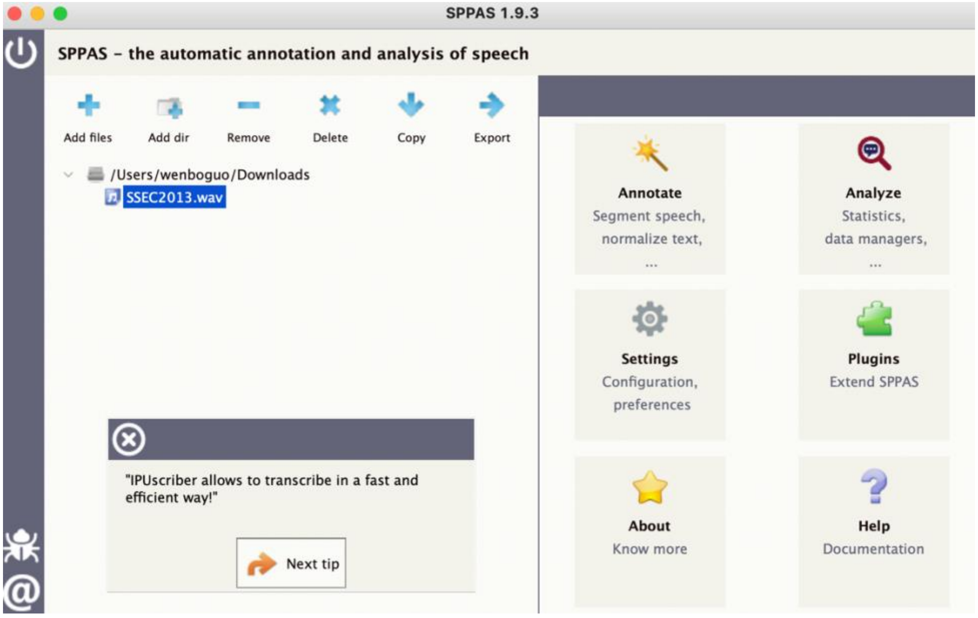


Figure 3 Interface of SPPAS – adding file

- Step 5: Continue on SPPAS. Click "Annotate" icon on the left side of the interface, which will pop-up a menu from which choose "IPUs Segmentation". A "Procedure Outcome Report" would appear on the screen, click close to quit the report. A file named SSEC2013.xra would appear below SSEC2013.wav. Click the left-pointing arrow on the "Annotate" menu to return to the panel.
- Step 6: Continue on SPPAS. Select both files on the left and click "Analyse" then choose "IPUscriber" to enter the window for transcription (see Figure 4 below). Use the panel at the bottom to replay the speech segments if needed. After transcription is compete, click "save". Click the "◌" sign then click "yes".

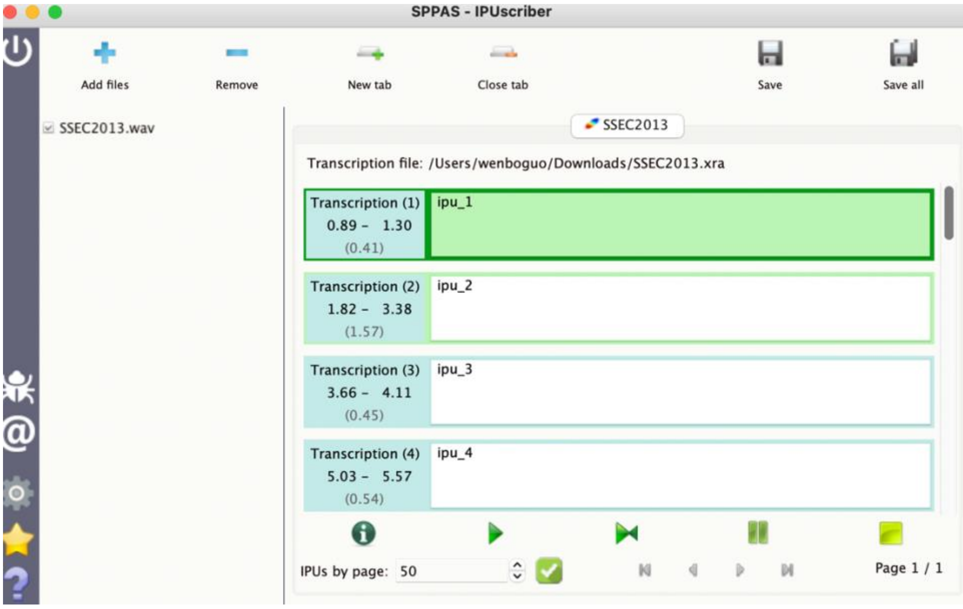


Figure 4 Interface of transcription window in SPPAS

- Step 7: Continue on SPPAS. Select SSEC2013.xra and click “export”. Choose the .Textgrid format. Save the exported file in the SSEC2013 folder.
- Step 7: Open ELAN. From the interface, click "file" then click "new". Click “Add Media File” then choose the SSEC2013.wav. Click “ok”. Click “file” and choose “save” it the SSEC2013 folder as SSEC2013.eaf – this is the corpus file. The sound spectrum of the source speech would appear in ELAN.
- Step 8: Continue on ELAN. Click “file” and choose “import”. From the menu, select “Praat TextGrid file”. Click “browse” and locate the SSEC2013 folder to find and add SSEC2013.TextGrid. Click “ok”. A “Select coding” menu will pop up. Choose “UTF-8” then click “Ok”. On the next menu, click “next” then “finish”. The transcribed speech is imported in ELAN in a tier (see Figure 5 below).

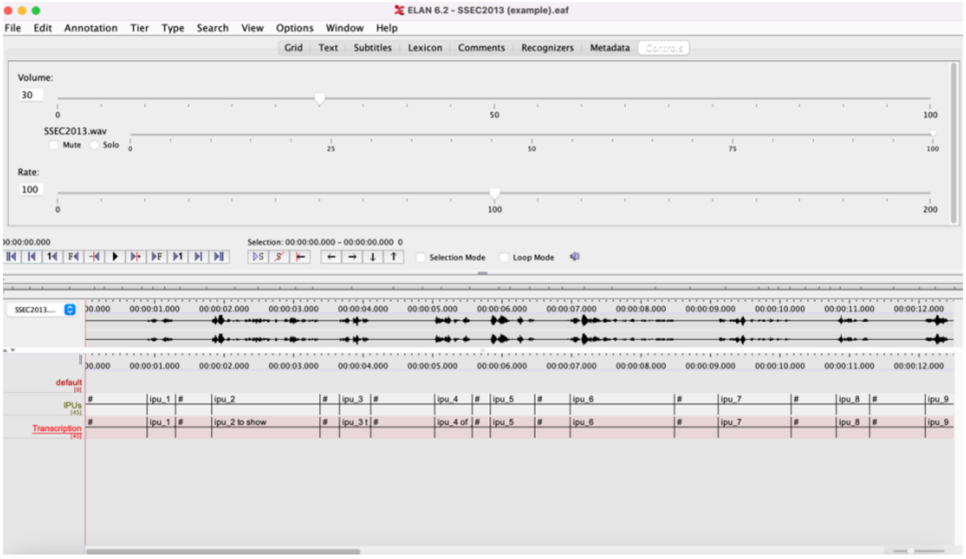


Figure 5 ELAN interface – after tier of transcription was added

- Step 9: Continue on ELAN. Click “Tier” then choose “Tokenise tier”. Select “Transcription” as the source tier then click “Create new tier”. Click “start”. A new window will appear, name the tier “tokenised source speech”. click “add” then “close”. Click “start” then “close”. A tokenised source speech named is added as another tier in the corpus.
- Step 10: Continue on ELAN. Double click “tokenised source speech” to select the tier then click “file. Hover the cursor on “export as” then choose Praat TextGrid. On the next window (see Figure 6 below), tick “tokenised source speech” then click “ok”. Name the new file “SSEC2013Praat” and save it in the SSEC2013 folder. A “Select coding” menu will pop up. Choose “UTF-8” then click “Ok”

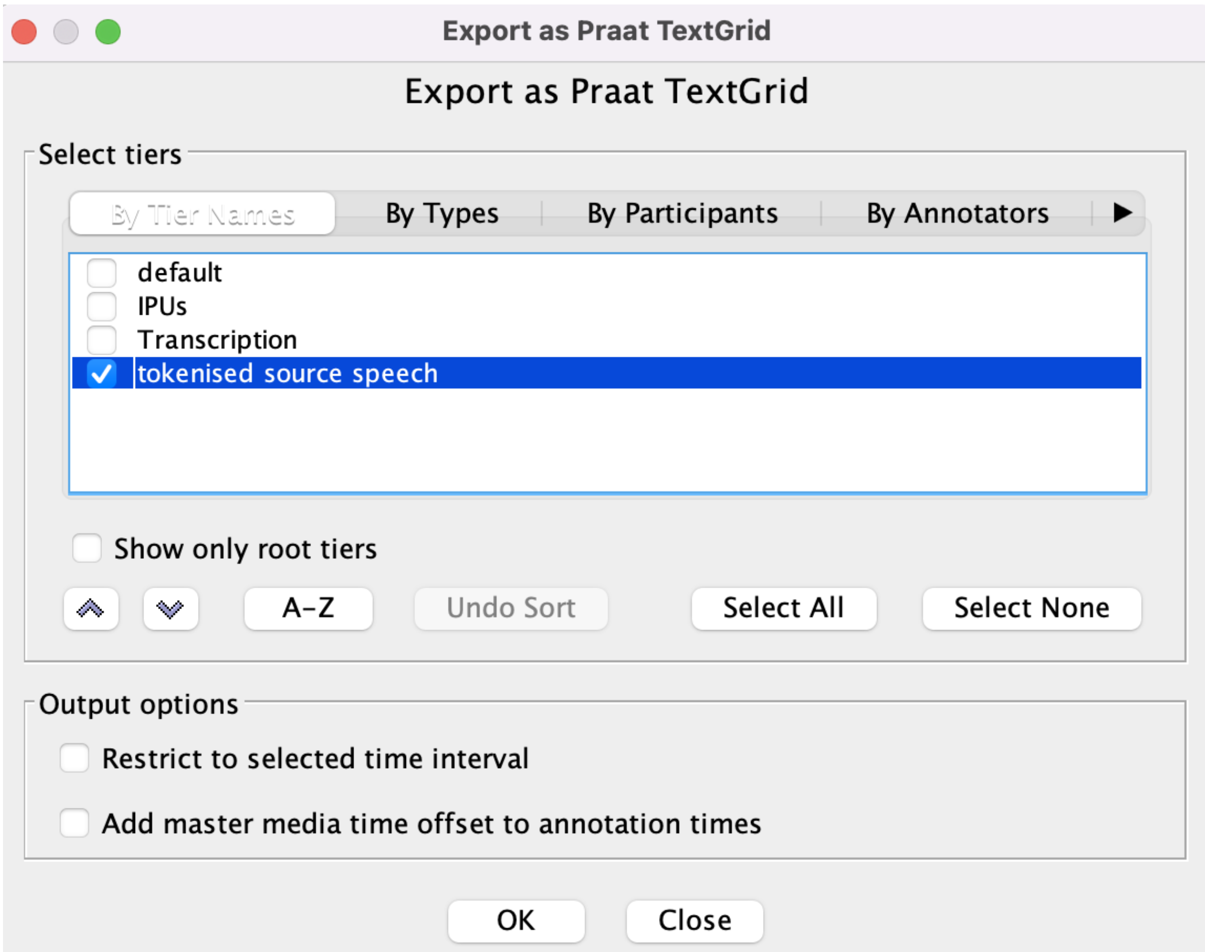


Figure 6 ELAN interface – tokenise speech

- Step 11: Open Praat. Two windows will open, one is "Praat Objects", and the other is "Praat Picture". Close "Praat Picture". Click "open" and select "read from file". Choose SSEC2013Praat.TextGrid saved in the SSEC2013 folder. Click "open" and select "read from file". Choose SSEC2013.wav from the same folder. Keep both files selected and click "view & edit".
- Step 12: Continue on Praat. The editing window can be divided to three sections (See Figure 7 below). The first section indicates the audio frequency of the source speech. The second section is the tokenised source speech that needs manual adjustment. Click on any word to edit. The third section indicate the duration of the speech. To play the audio segment of a word, click select the word then click the duration box right under the word. Drag the blue lines to secure an audio segment that is match the word. Mark pauses with a hashtag sign. Once all tokenised words have been adjusted, save the file and exit Praat.

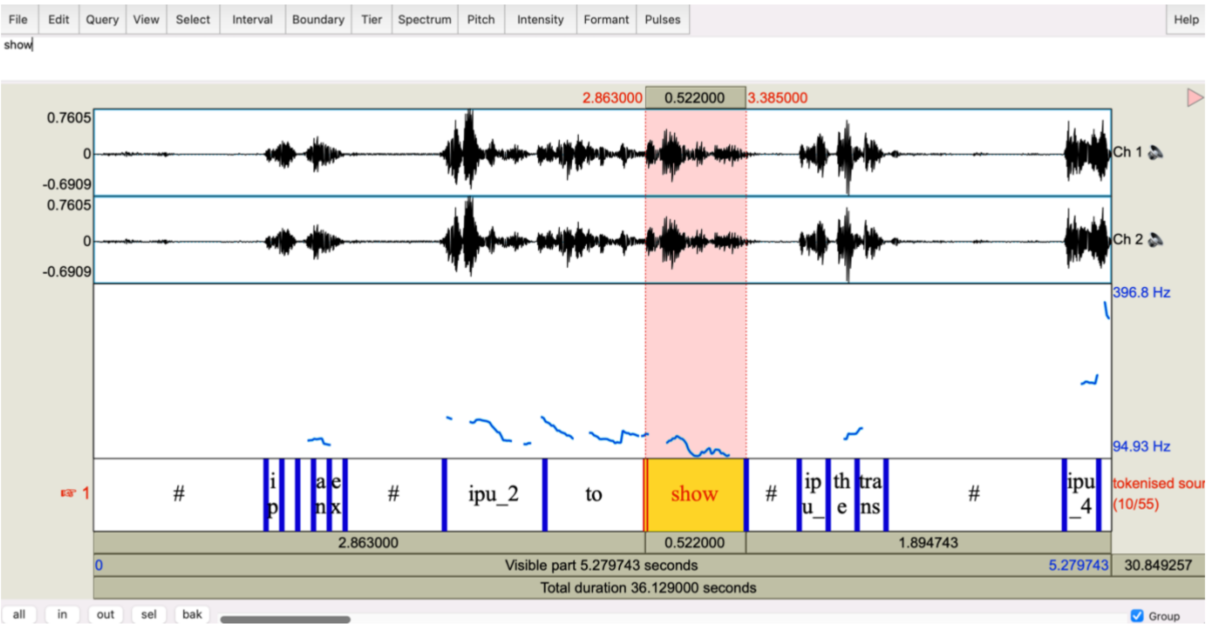


Figure 7 Interface of Praat – manually tokenise the source speech

- Step 13: Open SSEC2013.eaf in ELAN. Import SSEC2013Praat.TextGrid as a new tier. This the last step of building the corpus.

3.1.2.4 Corpus processing and coding

3.1.2.4.1 Corpus processing

Scholars in corpus linguistics have accentuated that building a corpus does not suffice to make it functional and it is essential to process a corpus with tools to generate data (Atkins *et al*., 1992; Rozas and Barcala, 2020; Knight, Morris, Arman, Needs, and Rees, 2021). To satisfy the demands of this corpus study and test the hypothesis that low word-frequency is one of the main GA inducers, the word frequency of each word from all source speech was obtained.

The main challenge in obtaining word frequency was to choose a suitable database. Studies investigating word frequency have mainly resorted to two famous databases: British National Corpus (BNC, Kilgarriff, 2006) and CELEX (Bayern, Pipenbrock, and Gulikers, 1995). Although both databases assembled a large number of words (100 million for BNC, 17.9 million for CELEX), they are not chosen for this study for two considerations. First, two databases sourced words mainly from written language. As a spoken corpus, considering the vast differences spoken and written language regarding formality and function (Brown, 1978; Brown, Gillian, Brown, and Yule, 1983; Cienki, 2015), it would be warranted to argue that databases composed of written language would not accurately predict the effect of word frequency in a spoken language task. Scholars suggested that subtitle-based word frequencies are more suitable in predicting frequency effects (Brysbaert, Buchmeier, Conrad, Jacobs, Bölte, and Böhl, 2011). Second, both BNC and CELEX collected from documents produced in the 1990s, and therefore their word frequencies might not truly reflect the usage of English today.

The database used to obtain word frequencies in this corpus was SUBTLEX-UK (Heuven *et al*., 2014) that assembled over 200 million words from the subtitles from British English TV programmes broadcasted in nine BBC channels between 2010 and 2012. Another merit of SUBTLEX-UK is that it improved the traditionally standarised frequency measure – frequency per million words (fpmw), which, according to Heuven *et al*., does not always correctly reflect the WFE because it is susceptible to sample size. Instead, the authors adopted a nominal Zipf scale. According to Heuven *et al*. (2014, p. 1179) a Zipf value "equals log10 (frequency per million words) +3", meaning that "a Zipf value of 1 correpsonds to words with frequencies of 1 per 10 million words". According to the scale, words with Zipf values between 1 and 3 are LFWs. Since its debut, the database has gained growing popularity in corpus studies that concern word frequency (e.g., Chen, Dong, and Yu, 2018; Abasq, Dabouis, Fournier, and Girard, 2019). Based on its merits and precedent applications, SUBTLEX-UK was deemed appropriate for this corpus study.

3.1.2.4.2 Definition and protocol of Annotation

In a corpus study, coding, or adding mark-ups, is the stage when the corpus begins to function (Atkins *et al*., 1992). Scholars in corpus studies have accentuated the significance of developing clear definitions and protocol of mark-ups (Upton and Cohen, 2009; Chang and Huang, 2015). For the current corpus study, in pursuing the general research aim that is to code and explore interpreters' GA during CI active listening and notetaking, two mark-ups were initially developed: GA+NT and GA-NT, with the former indicating interpreters' behaviour of looking away from notepad while writing and the latter suggesting interpreters' behaviour of looking away from notepad while not writing. The mark-ups were developed via the Controlled Vocabulary function in ELAN. Table 4 as follow is a summary of the codes and their corresponding descriptions.

Table 4 Mark-ups and Definitions

Code	***Indication***	***Description***

GA+NT	GA with notetaking	It describes the moment when an interpreter looked away from notepad while writing.
GA-NT	GA without notetaking	It describes the moment when an interpreter looked away from notepad while not writing. In other words, the interpreter looked way from notepad to focus on listening.

An annotation protocol was designed for coding GA in the corpus. A key issue to address when developing the protocol was how to decide the onset and completion of GA as eye movements can be subtle (Bayliss *et al*., 2006). For this study, the onset of GA is defined to be the moment when an interpreter demonstrated the readiness to look way from the notepad, which either resulted in a quick move of eyeball (subtle GA) or a turning of head (overt GA). The completion of GA is defined to be the moment when an interpreter re-fixated on the notepad. However, in rare situations, exceptions were applied to the eye movements that met the above description (mostly subtle GA) but were suspected to be task irrelevant. For instance, when an eye movement was believed to be conducted to alleviate the discomfort in the eye, the movement should not be coded as GA.

3.1.4.2.4 Coding Procedure

Once the definitions and protocol of annotations were developed, corpus was ready for coding. Using SSEC2013.eaf mentioned in 3.1.2.3 as an example, the following offers a step-by-step coding procedure:

- Step 1: Open SSEC2013.eaf. Add a new tier to the corpus and assign the tier name to be Interpreter X, with “X” being a number that indicates the order of coding (e.g., Interpreter 2);
- Step 2: Open a video in QuickTime, and set the playback speed to be half of the original speed to enhance the accuracy of detecting the onset and completion of GA;
- Step 3: When GA is observed (during active listening and notetaking only), find the corresponding onset and completion locations in the source speech in ELAN. These locations are usually a phonetic unit of a word. The source speech can be zoomed-in so that locating a desired phonetic unit is more straightforward. In the tier that is under coding, double click the location of the onset then hold and drag the cursor to the location where the GA is completed. Double click the selected area and choose annotation from the controlled vocabulary that reflects the type of GA (i.e., whether the GA was accompanied by notetaking or listening-only). Repeat this step until the end of the final speech segment.

3.1.4.2.4 Inter-rater Reliability Check

Inter-rater reliability checks are necessary in corpus-based studies to ensure the accuracy of coding (Biber, 2007; Upton and Cohen, 2009). To abide by the standard, an inter-rater (female, 27) was recruited to check the reliability of coding. The coder is a German native speaker studying as a PhD candidate at Newcastle University. Before being admitted to the programme, she obtained a score of 8 in an IELTS (International English Language Testing System) test, meaning that she is a “very good user” of English. Also, has experience in coding corpus as she had completed a corpus study.

In total, 4 out of 30 videos were randomly chosen for coding. Before the task, the coder went through a training session during which the definitions and protocol of annotations were provided in written forms, which was followed by explanation and clarification for details. During the session, the coder practiced adding and adjusting annotations in ELAN. There was little discrepancy regarding the occurrence and the type of GA, meaning that the definitions and protocol resulted in good understanding. There were some discrepancies regarding the length of some GA, but the differences are minimal and most of the differences were reconciled during the post-coding discussion. The inter-rater reliability check rendered a 97% agreement, with the disagreement reconciled through discussion.

3.1.5 Data Collection

3.1.5.1 Collecting Corpus Data

Collecting data from the corpus was achieved via two steps. Owing to the user-friendly search-and-export function of ELAN, mark-ups for each interpreter were firstly found (See Figure 8 for an example) and exported as a tab-delimited file which contains information regarding counts and length of GA. The information was then manually copied and organised in Microsoft Excel. The attributes include GA type, total GA counts, total GA length, total GA counts per 100 source speech words, total GA counts per 100 source speech seconds, and total GA length to total speech length. Mark-ups for all interpreters from the same year were organised in the same file to facilitate analysis.

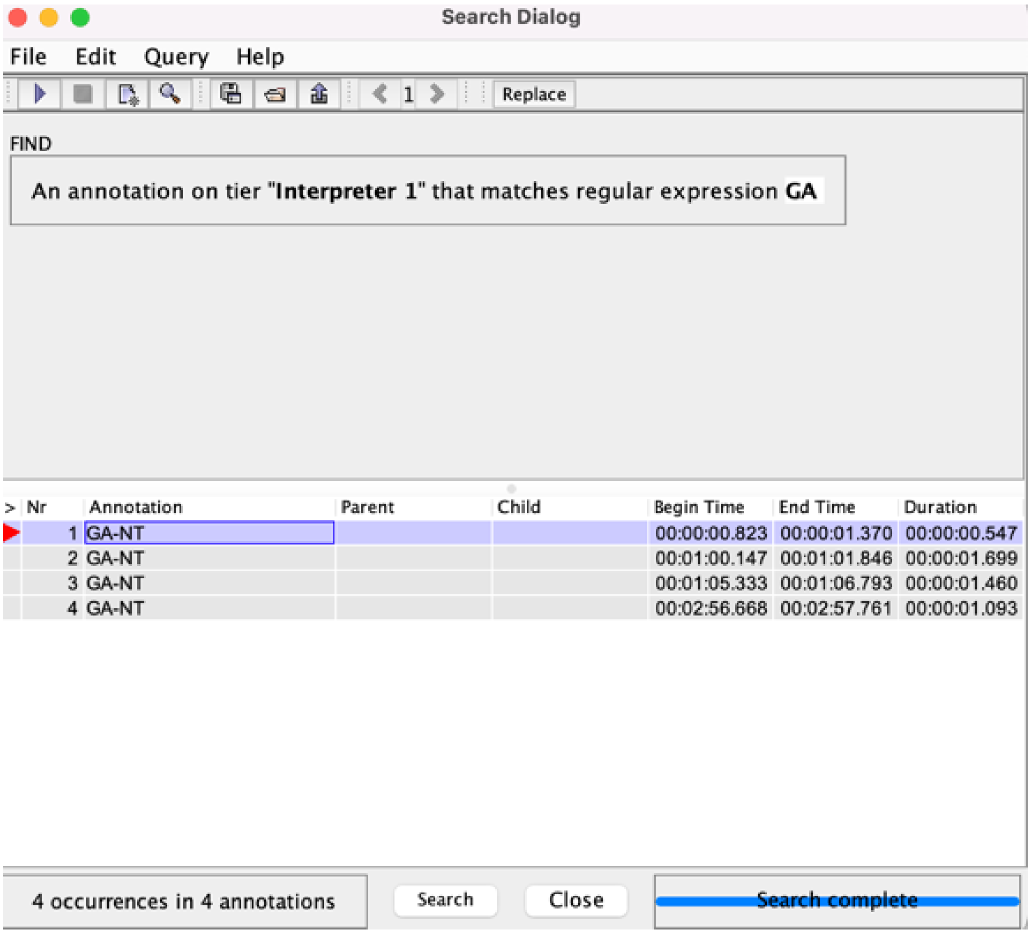


Figure 8 Interface of ELAN search box

3.1.5.2 Identifying of Textual Features

The existing literature on GA suggested that the behaviour would be performed in response to high cognitive load (Glenberg *et al*., 1998; Doherty-Sneddon *et al*., 2002; Doherty-Sneddon and Phelps, 2005), implying that GA is triggered by activities that are cognitively expensive. In interpreting studies, a number of elements, such as numbers, are deemed "problem triggers" as they could lead to WM overload (Gile, 2009). However, linguistically speaking, it remains unknown what elements would trigger GA. To narrow down the potential GA triggers in CI textual features of the speech contents immediately preceding each GA were identified and organised in an Excel spreadsheet where GA attributes are organised. See Table 3.5 in section 3.1.8 for an example.

3.1.5.3 Scoring Sentence Difficulty

To recap, a part of the hypothesis developed for RQ3 is that LFWs or subordinate clause or both would cause more and longer GA. To test the hypothesis, the difficulty of each sentence of the source speech needs to be measured. Some studies designed scoring systems to investigate sentence complexity (e.g., Flick, 1977; Ran, 2021). In the context for this corpus study, a scoring system was developed to measure the difficulty of each sentence of the source speech. According to the system, each subordinate clause or a word with a zipf value less than 3 would denote a score as 1, respectively. Sentences that do not contain subordinated clauses or LFWs would have a score of 0. See the following three sentences as examples:

(1) Does it make any difference to your class calculator ***when*** *you hear "****how do you do****" versus "****Pleased to meet you****"?* (Difficulty score: 3)
(2) And we want to tell our grandchildren ***that*** *there was a terrible time in history* ***where*** *up to a third of the children had brains and bodies* ***that*** *were* ***<u>stunted</u>****, but that exists no more.* (Difficulty score: 4)
(3) I read an article in the Wall Street Journal today and I can't think of a better piece of news to begin my talk with. (Difficulty score: 0)

Sentence (1) was given a difficulty score of 3 because it contains 3 subordinate clauses, with the first one introduced by ***when*** and the rest two without any introducer. Sentence (2) reached a higher score with three subordinate clauses introduced by ***that***, ***where***, and ***that***, respectively, and a LFW (***stunted***). Sentence (3) scored 0 because it contains neither subordinate clause nor LFW.

Using the system, a difficulty score was yielded for each source speech sentence and organised in a spreadsheet for later analysis. To show an example, the following Table 5 listed a few sentences from the source speech used in 2016:

Table 5 Difficulty score for each source speech sentences

Sentence Number	***Sentence content***	***Difficulty Score***
1	I read an article in the Wall Street Journal today, and I can't think of a better piece of news to begin my talk with.	0
2	Two Italian economists compared data on ***Florentine*** taxpayers from 1427 against tax data from 2011 and found about 900 surnames still present in Florence.	1
3	It would appear ***that*** the wealthiest families in Florence today are descended from the wealthiest families of Florence nearly 600 years ago.	1

4	I later found out ***that*** descendants of Japan's samurai remain elites 140 years after their ancestors gave up their swords.	1
5	Even the communist party in China failed to drive social mobility.	0
6	It implies **that** even universities are **stratified** too.	2

Although the word frequency values were retrieved from a database, meaning that the difficulty scores imposed by LFWs were not results from subjective scoring, identifying subordinate clauses was achieved via subjective judgement. Therefore, an inter-rater (female, 28) was called in to check the reliability of the scores. The inter-rater is a native Chinese speaker studying in Newcastle University as a PhD candidate. She obtained a master's degree in Translation and Interpreting from Newcastle University in 2017 and has been working as a freelance interpreter while studying in the same institute as a PhD candidate. One of the four source speech was randomly chosen for the task. An initial 98% of agreement was reached, and all discrepancies were reconciled.

3.1.6 Data analysis

3.1.6.1 Descriptive analysis

As previously discussed, this corpus study adopted a mixed-methods design where quantitative data was yielded via coding the corpus and identifying textual features of the source text preceding each GA. Coding a corpus usually results in abundant and messy data (Stuart *et al.*, 2016). For this corpus study, the volume of quantitative data was further enriched by the identified textual features immediately preceding each GA for each interpreter. To organise the data in a way that facilitates straightforward reviews and analyses, a master spreadsheet listing all measurements for each interpreter was created for each year. To show an example, the table below is extracted from the master spreadsheet. Based on the master spreadsheet, data were organised to produce tables and diagrams that facilitate results reporting in 3.2.

Table 6 Master sheet for all GA

Interpreter	GA Type	Total GA Counts	Length per GA (second)	Total GA Length (second)	GA Counts/100 Words	GA Counts/100 Seconds	GA Length/Speech Length (proportion ratio)	Textual Features Preceding and During GA
1	GA-NT	4	0.55	4.8	0.76	1.80	2%	NA[1]
	GA-NT		1.7					Complex Sentence Structure
	GA-NT		1.46					Complex Sentence Structure
	GA-NT		1.09					Number + LFW
2	GA+NT	7	0.88	4.59	1.33	2.07	2%	Complex Sentence Structure + Passive voice
	GA-NT		0.32					Complex Sentence Structure
	GA-NT		0.84					Complex Sentence Structure + Passive voice
	GA-NT		0.78					Complex Sentence Structure + LFW
	GA-NT		0.58					Complex Sentence Structure + LFW
	GA-NT		0.8					Complex Sentence Structure
	GA+NT		0.39					LFW

3.1.6.2 Inferential Analysis

As previously mentioned, the overall purpose of the study and the quantifiable data led to the adoption of a quantitative design. For RQ1, a question that breaks the ground for the thesis by asking if interpreters would perform GA at all, was adequately answered via descriptive analysis. For RQ2 and 3, the two questions that lead to zoomed-in inspection of GA, required further efforts once descriptive analyses were completed. Specifically, for RQ2, inferential analyses were performed to test whether the difference between the two types of GA (GA+NT vs GA-NT) reached statistical significance. For RQ3, inferential analyses were conducted to explore two aspects: to test if GA was performed significantly more frequently after textual features that could potentially increase cognitive load were present in the source text; to examine whether GA sentences (sentences where interpreters performed GA) are significantly more difficult than non-GA sentences (sentences where interpreters did not perform GA). According to scholars, comparing data formed in two groups would be robustly achieved through paired sample *t*-test, provided the assumption of normality is verified, or via Wilcoxon signed-rank test if the normality check did not support normal distribution of data

[1] The textual feature is non-applicable as the text preceding the GA was not a part of the source speech but audio introduction of the exam.

(Field, 2013; Gerald, 2018). Therefore, the assumption of normality for data obtained for both RQs were tested prior to finalising the selection of inferential tests.

3.2 Results

3.2.1 Mark-ups from the corpus (RQ1)

3.2.1.1 Overall description

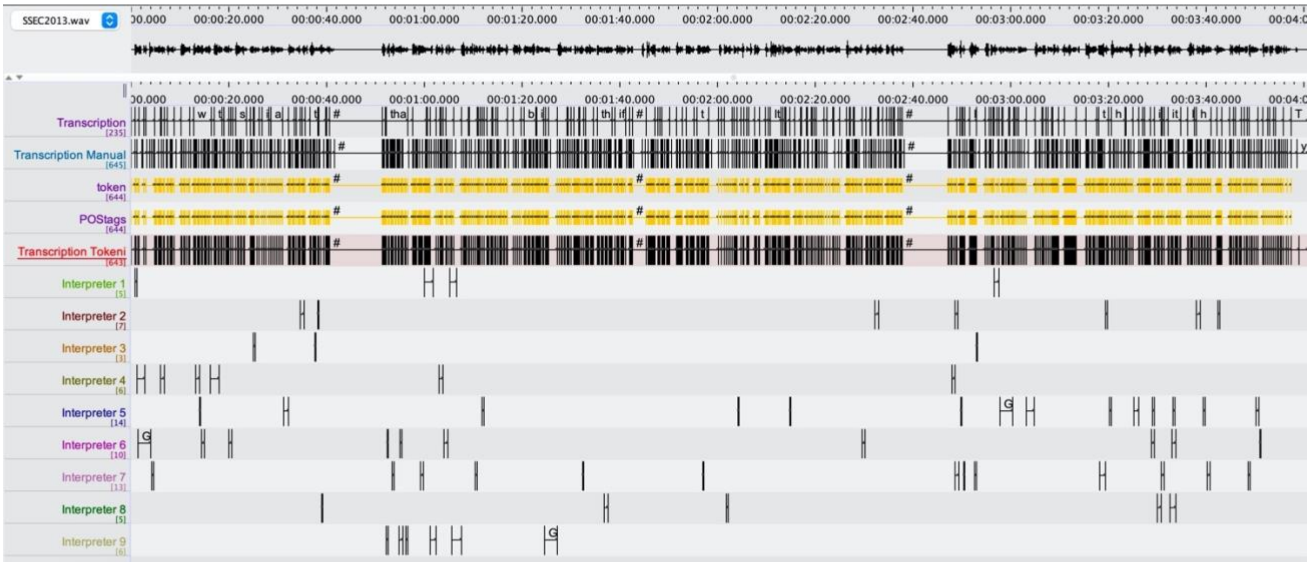


Figure 9 ELAN interface after coding

Figure 9 represents an example of the interface of ELAN after coding. In the example, there are 9 independent tiers with GA mark-ups, with each layer indicating one interpreter. At first glance, each horizontal layer is dotted with mark-ups, with each of which representing an occurrence of GA. In other words, the interpreter averted gaze from the notepad at that moment during CI active listening and notetaking. The features observed from the example are also seen in other cohorts. An initial observation of the mark-ups across the years revealed that all interpreters performed GA at some points during the course of notetaking and active listening.

Table 7 Overall GA data

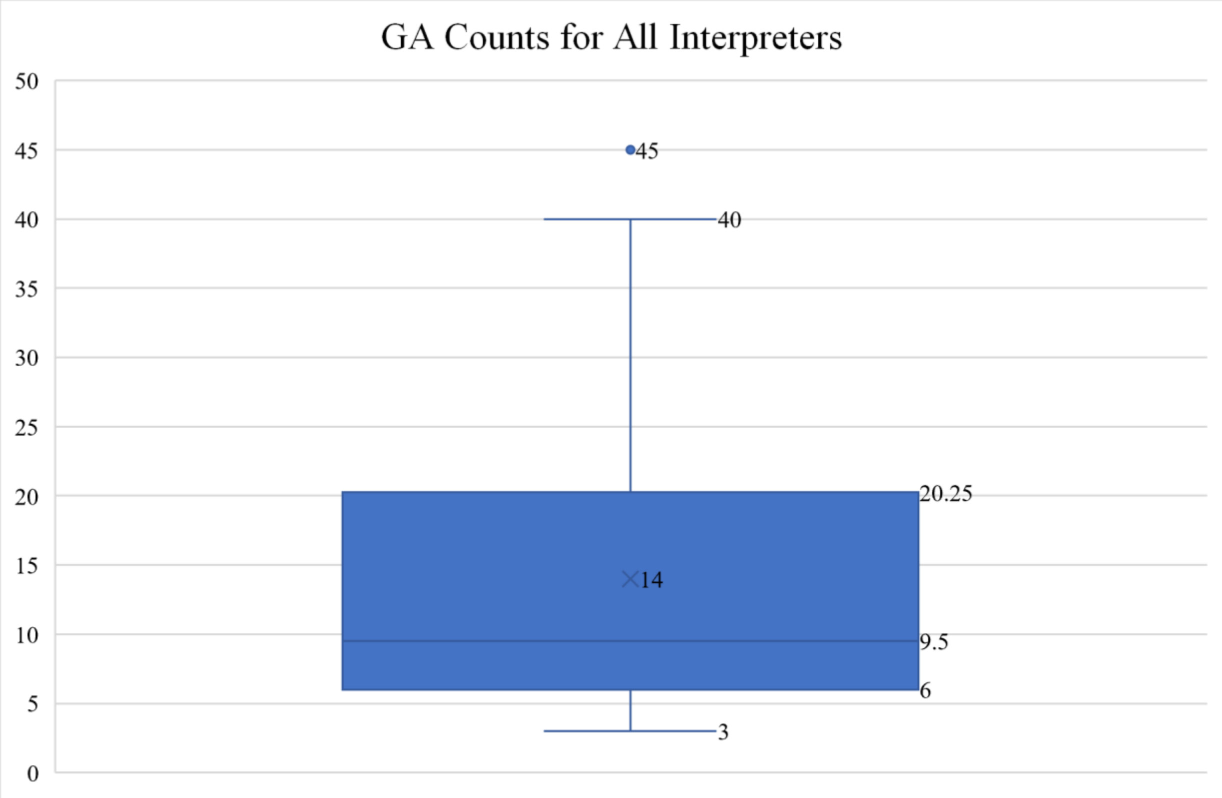
Year	***Number of Participants***	***Total GA Counts***	***Counts per Participant***	***Total GA Length (Seconds)***	***Average GA Length (Seconds)***
2013	9	69	7.67	53.34	0.77
2014	5	56	11.20	73.1	1.31

2015	7	106	15.43	158.7	1.48
2016	9	188	17.44	196.9	1.05
Total	30	420	N/A	482.04	N/A

Table 7 above summarises the overall information regarding the number of interpreters and overall GA data across four years. In total, 30 interpreters performed 420 times of GA, with a total time spent in GA lasting 482.04 seconds. On average, despite the source speech variances, each interpreter spent roughly 16 seconds to perform 14 times of GA. Considering that the four source speech contribute to a total length of 866 seconds (see table 3.1), the 420 times of GA means that GA occurred at a frequency of roughly 28 times of GA per minute.

Another overall feature in the ELAN interfaces is the considerable discrepancy regarding the occurrence of GA, as some of the marks are much more scattered than others, indicating strong individual differences.

Two boxplots (Figure 10) below are plotted to display the distribution of GA counts and average GA length, respectively, for all interpreters in the corpus study. In terms of counts, the median score of GA counts is 9.5, with the lowest number being 3 and the highest being 45. The interquartile range is between 6 (lower quartile (Q1)) and 20.25 (upper quartile, (Q3)), suggesting that about 50% of the participants performed between 6 and 20.25 times of GA. Overall, it can be seen that the data is positively skewed with an outlier of 45. Regarding average GA length, the median score is 0.99 seconds per GA, with the shortest GA lasting 0.3 seconds and the longest lasting 1.84 seconds. The interquartile range is between 0.71 (Q1) and 1.35 seconds (Q20), indicating that 50% of GA across the four years lasted between 0.71 to 1.35 seconds. Compared with GA counts, an initial observation of the box would suggest that the data on average GA length is more symmetrically distributed. Despite the different data distribution, both box plots indicate strong individual differences regarding both measures, a pattern echoing with previous eye-tracking studies (Rogers *et al*., 2018; Peterson, Lin, Zaun, and Kanwisher, 2016) where evident individual differences in gazing behaviour were reported.



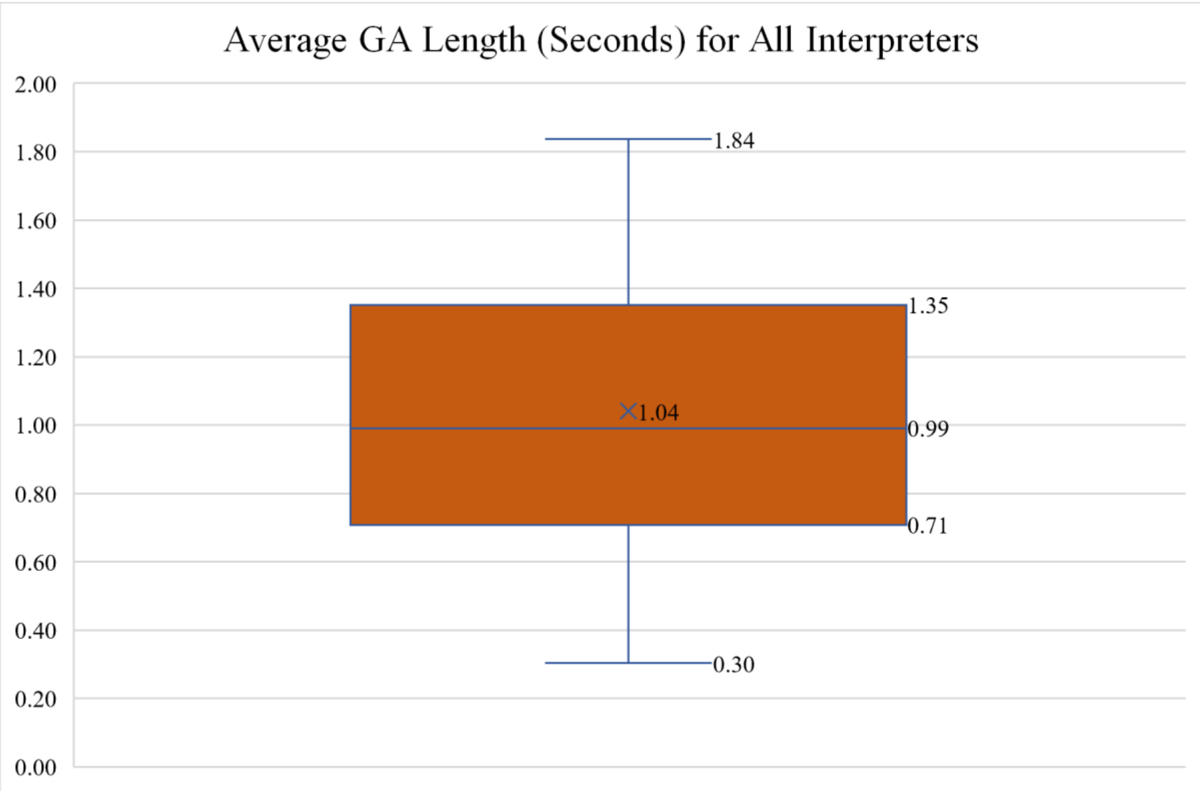


Figure 10 GA counts and average GA length for all interpreters

The overall results from coding GA validated the hypothesis of RQ1, which asks whether interpreters would avert their gaze from notepads during CI. Put simply, the results showed that it is common for interpreters to avert their gaze from notepads during active listening and notetaking, but the frequency and length of GA are subject to considerable individual differences. As discussed in 3.1, the coding of the corpus constituted a master task that generated data and indications for analysis. Therefore, the finding that GA is widely shared across the participants yielded overarching significance to this thesis as it paved the avenue for investigating other RQs and the intricate layers of GA.

3.2.1.2 GA frequency and proportion

As mentioned in 3.1, several measurements of GA were explored. Specifically, GA was measured in terms of total GA counts, GA counts/ 100 words, GA counts/100 seconds, total GA length, and GA proportion (time spent in GA divided by length of the source speech). Considering the aim of revealing the intricate layers of GA, it would be warranted to provide a zoomed-in view by separately reporting results obtained from each year.

3.2.1.2.1 GA Data from the 2013 Cohort

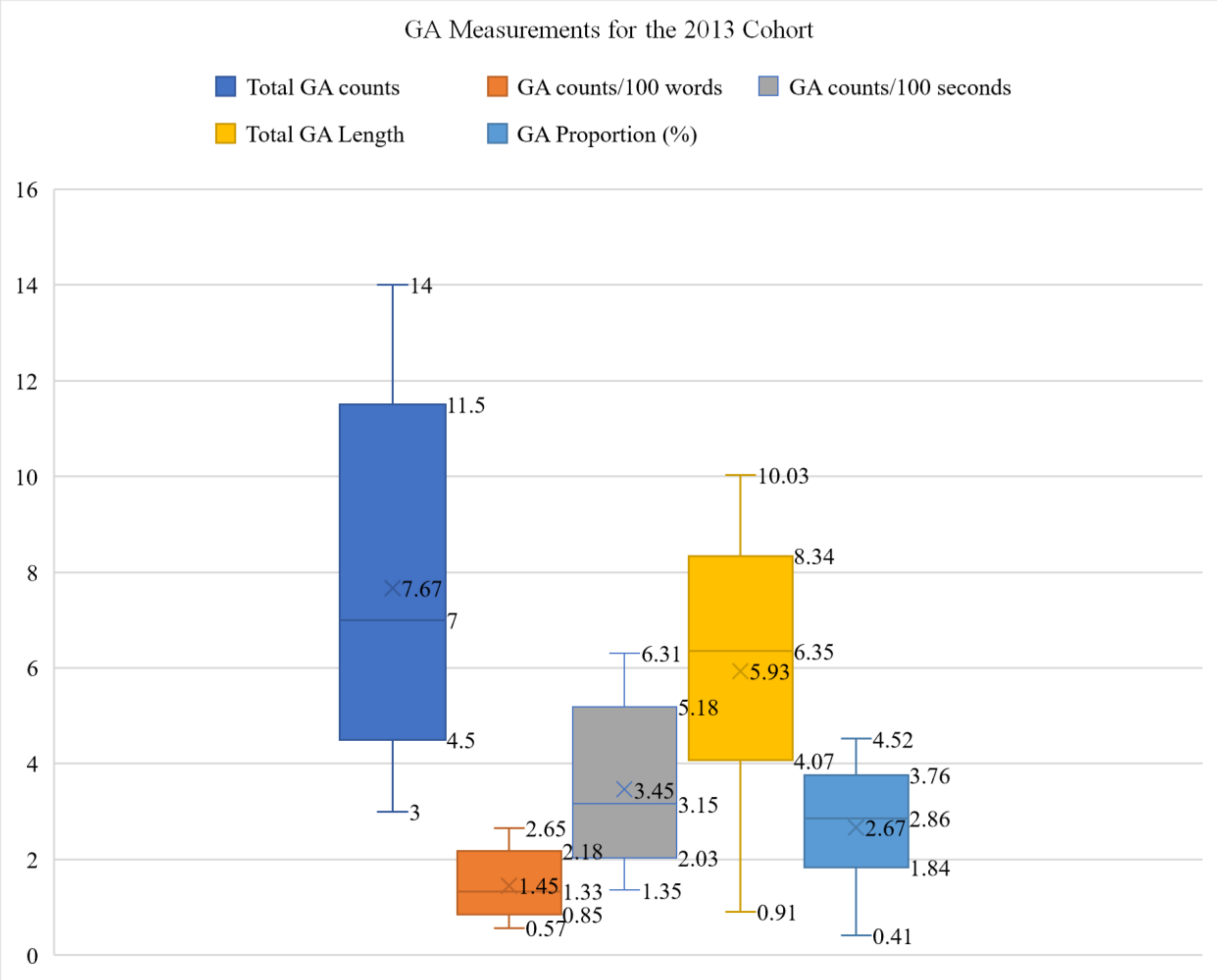


Figure 11 Box plot for GA measurement for the 2013 cohort

Figure 11 above is plotted to display the distribution of GA measures pertaining to frequency and proportion based on the master sheet (see Appendix G) for the 2013 speech. In terms of total GA counts, the median score of total GA counts is 7, with an interquartile range between 4.5 (Q1) and 11.5 (Q3), indicating that half of the interpreters whose videos were coded performed between 4.5 and 11.5 times of GA. On average, each interpreter in 2013 conducted 7.67 times of GA (*max* = 14, *min* = 3). Regarding total GA length, the median score is 6.35 seconds, with an interquartile range between 4.07 and 8.34 seconds. On average, each interpreter in 2013 spent 5.93 seconds looking away from their notepads (*max* = 10.3, *min* = 0.91). For GA counts per every 100 words from the source speech, the median score is 1.45, with a Q1 at 0.85 and a Q3 at 2.18. On average, 1.45 times of GA were coded for every 100 words uttered by the speaker (*max* = 2.65, *min* = 0.57). The median number of GA coded every 100 seconds of the source speech is 3.45, with Q1 at 2.03 and Q3 at 5.18. On average, each interpreter performed 3.45 times of GA during each 100 seconds of the source speech (*max* = 6.31, *min* = 1.35). Finally, regarding the proportion of time spent in GA to speech length, the median score is 2.67%. The interquartile range is between 1.84% (Q1) and 3.76% (Q3), suggesting that half of the interpreters spent 4.08 to 8.35 seconds averting gaze from notepads. On average, 2.67% of the time (or 5.93 seconds) was invested in GA (*max* = 4.52%, *min* = 0.41%). Overall, the box plot does not indicate any outliers across the measures or distinctive skewness.

3.2.1.2.2 GA Data from the 2014 Cohort

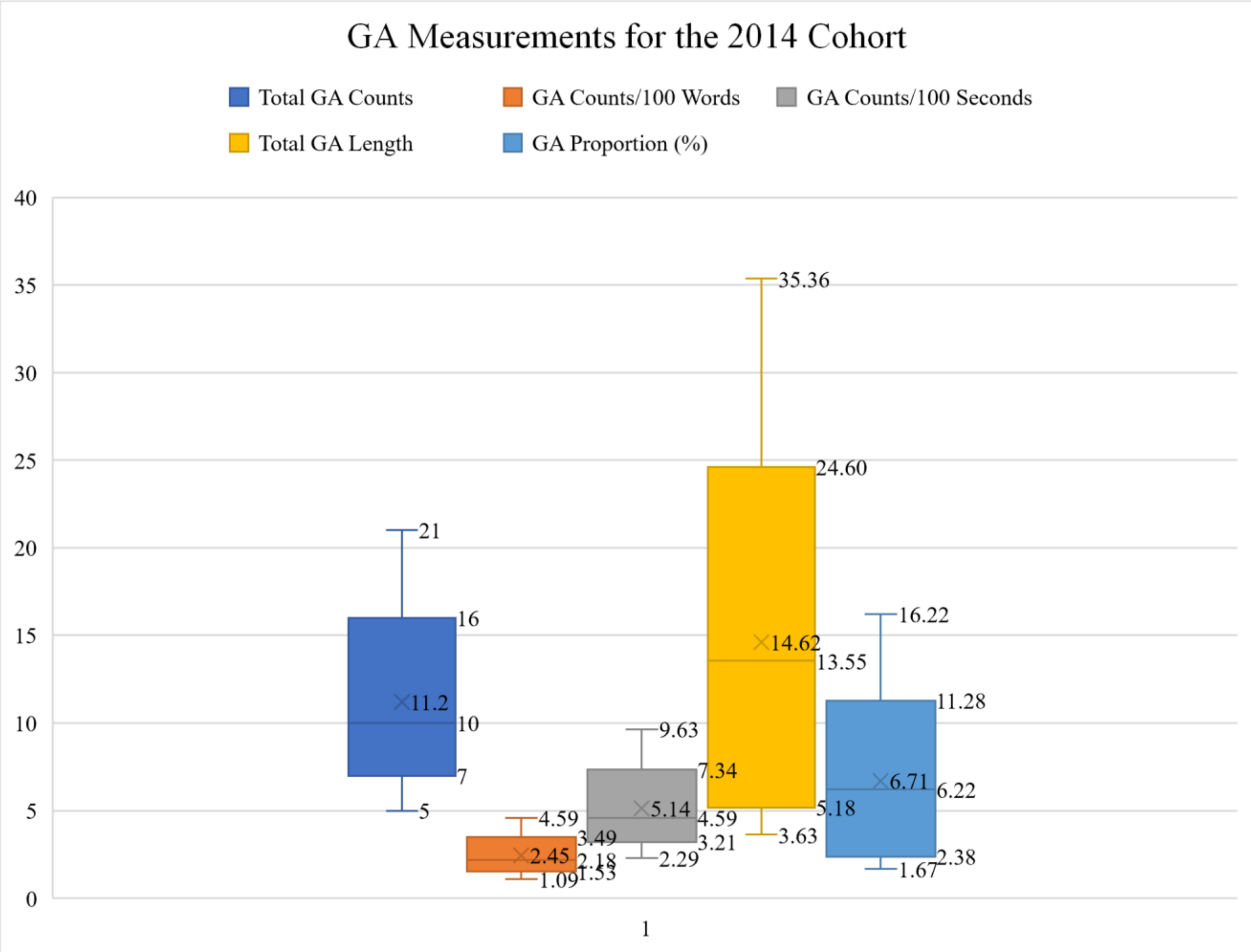


Figure 12 Box plot for GA measurements for the 2014 cohort

Figure 12 above is plotted to display the distribution of GA measurements for the 2014 speech based on the master sheet (See Appendix G). Regarding total GA counts, the median score is 10. The interquartile range is between 7 (Q1) and 16 (Q3), meaning that half of the interpreters from the 2014 cohort performed between 7 and 16 times of GA. On average, each interpreter in 2014 conducted 11.2 times of GA (*max* = 21, *min* = 5). The median score for the total GA length measure is 13.55 seconds, with an interquartile range between 5.18 and 24.6 seconds. On average, each interpreter in 2014 spent 14.62 seconds averting gaze from their notepads (*max* = 35.36, *min* = 3.63). For GA counts per every 100 words from the source speech, the median score is 2.18, with a Q1 at 1.53 and a Q3 at 3.49. On average, 2.45 times of GA were coded for every 100 words uttered by the speaker (*max* = 4.59, *min* = 1.09). The median number of GA coded every 100 seconds of the source speech is 4.59. The interquartile range is between 3.21 (Q1) and 7.34 (Q3). On average, each interpreter performed 5.14 times of GA during each 100 seconds of the source speech (*max* = 9.63, *min* = 2.29). Finally, in terms of the proportion of time spent in GA to speech length, the median score is 6.22%, with an interquartile range between 2.38% (Q1) and 11.28% (Q3), suggesting that half of the interpreters spent 5.19 to 25.77 seconds looking away from notepads. On average, 6.71% of the time (or 14.63 seconds) was invested in GA (*max* = 16.22%, *min* = 1.67%). Similar to the 2013 cohort, data coded from the 2014 cohort does not show dramatic skewness.

3.2.1.2.3 GA Data from the 2015 Cohort

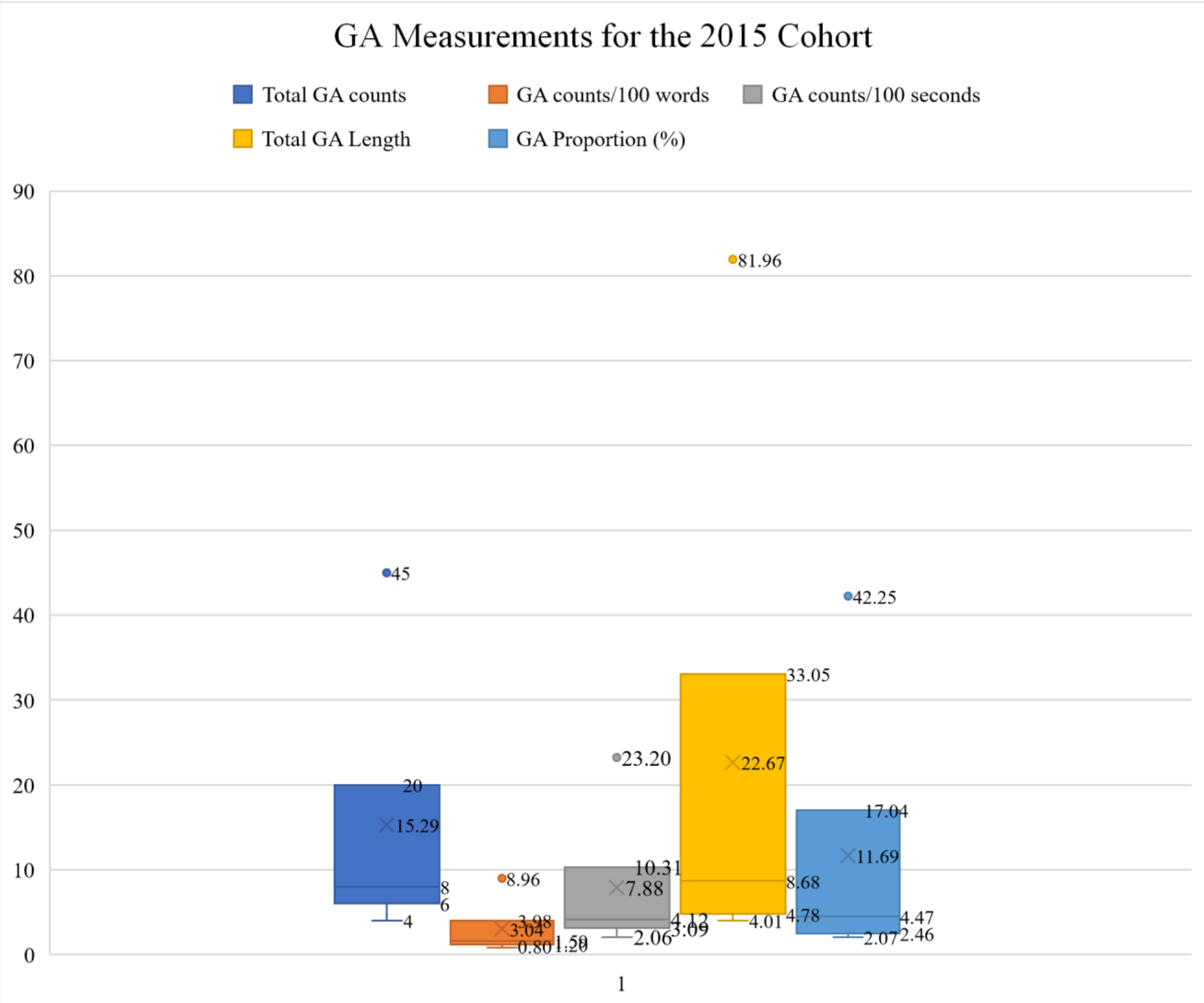


Figure 13 GA measurements for the 2015 cohort

The box plot (Figure 13) above illustrates the distribution of GA measurements from the 2015 cohort based on the master sheet (See Appendix G). In terms of total GA counts, the median score is 8, with an interquartile range between 6 (Q1) and 20 (Q3), suggesting that half of the interpreters averted gaze from notepads 6 to 20 times. On average, each interpreter from the 2015 cohort performed 15.29 times of GA (*max* = 45, *min* = 4). Regarding total GA length, the median score is 8.68 seconds, with an interquartile range between 4.78(Q1) and 33.05 seconds (Q3). On average, each interpreter invested 22.67 seconds in averting gaze from the notepad (*max* = 81.96, *min* = 4.01). The median for the score of GA counts per 100 words of the source speech is 1.5. The interquartile range is between 1.2 (Q1) and 3.98 (Q3). On average, 3.04 times of GA were coded for every 100 words uttered by the speaker (*max* = 8.96, *min* = 0.8). For GA coded every 100 seconds of the 2015 speech, the median score is 4.12, with an interquartile range between 3.09 (Q1) and 10.31 (Q3). The average amount of GA coded per 100 seconds is 4.12 (*max* = 23.20, *min* = 2.06). Finally, pertaining to GA proportion, the median value is 4.47%, with an interquartile range between 2.46% and 17.04%, suggesting that half of the interpreters spent 4.78 to 33.06 seconds conducting GA while processing the source speech (*max* = 42.25%, *min* = 2.07%). Compared with the 2013 and 2014 cohorts, it is evident that the distribution of GA data obtained from the 2015 cohort is positively skewed with the maximum scores being outliers located outside the left whisker.

3.2.1.2.4 GA Data from 2016

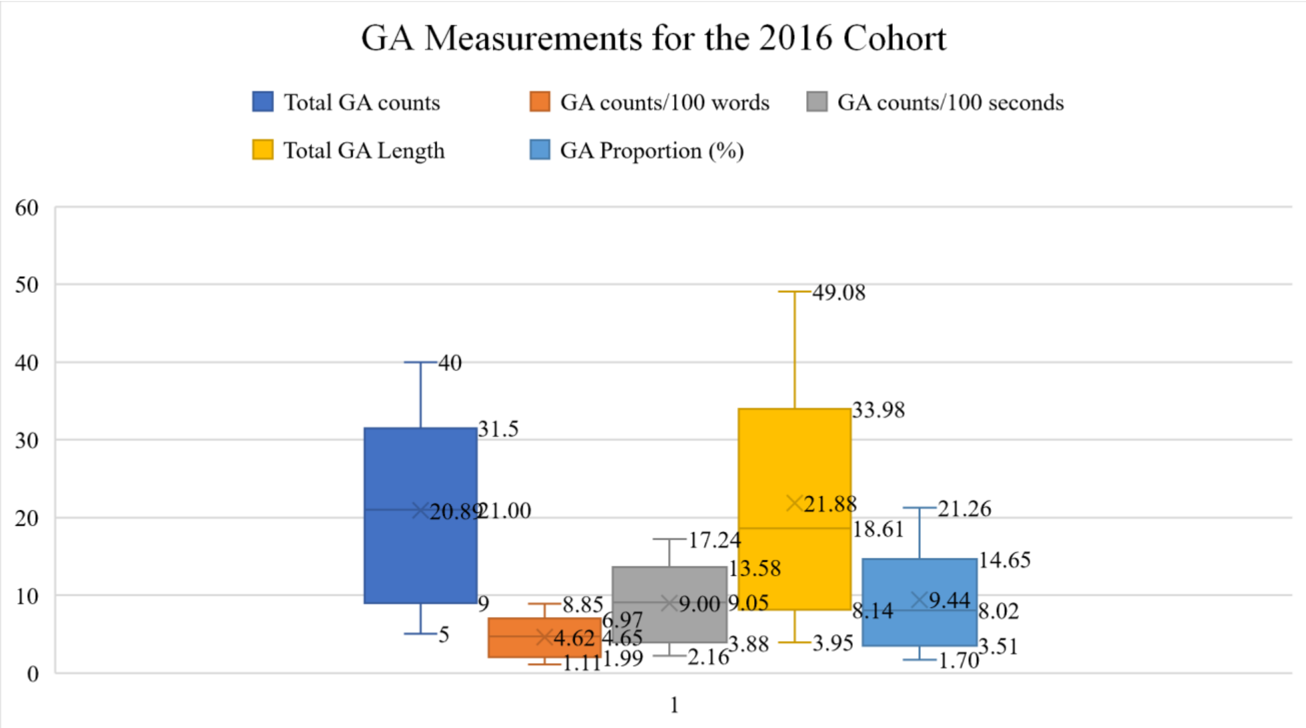


Figure 14 GA measurements for the 2016 cohort

Figure 14 above is plotted to display the distribution of GA measurements for the 2016 speech in accordance to the master sheet (See Appendix G). Regarding total GA counts, the median score is 21. The interquartile range is between 9 (Q1) and 31.5 (Q3), meaning that half of the interpreters from the 2016 cohort performed 9 to 31.5 times of GA. On average, each interpreter in 2016 conducted 20.89 times of GA (*max* = 40, *min* = 5). The median score for the total GA length measure is 18.61 seconds, with an interquartile range between 8.15 and 33.98 seconds. On average, each interpreter in 2016 spent 21.88 seconds averting gaze from their notepads (*max* = 49.08, *min* = 3.95). For GA counts per every 100 words from the source speech, the median score is 4.65, with a Q1 at 1.99 and a Q3 at 6.97. On average, 4.62 times of GA were coded for every 100 words uttered by the speaker (*max* = 8.85, *min* = 1.11). The median number of GA coded every 100 seconds of the source speech is 9.05. The interquartile range is between 3.88 (Q1) and 13.58 (Q3). On average, each interpreter performed 9 times of GA during each 100 seconds of the source speech (*max* = 17.24, *min* = 2.16). Finally, in terms of the proportion of time spent in GA to speech length, the median score is 8.02%, with an interquartile range between 3.51% (Q1) and 14.65% (Q3), suggesting that half of the interpreters spent 8.14 to 34 seconds looking away from notepads. On average, 9.44% of the time (or 22.9 seconds) was invested in GA (*max* = 21.26%, *min* = 1.7%). Data coded from the 2015 cohort does not show outliers or dramatic skewness.

3.2.1.2.5 Summary of Key Features

3.2.1.3 GA Type (RQ2)

3.2.1.3.1 Overall Description

The second RQ investigated in this initial corpus study concerns interpreters' action of notetaking accompanying GA. To be specific, it asked whether interpreters would be engaged in writing down notes while looking away from the notepad. It is hypothesised each interpreter would exhibit with a mixture of GA that accompanied by notetaking and GA without notetaking. Owing to the Controlled Vocabulary function in ELAN, mark-ups for coding GA in the corpus were specifically designed to be GA+NT and GA-NT. To recap, GA+NT represents the action of averting gaze while writing down notes, whereas GA-NT describes the type of GA that is not accompanied by handwriting. A scatter plot (Figure 15) below illustrates the distribution of two types of GA among the interpreters across four years, with each dot representing an interpreter. The x-axis shows the counts of GA-NT, whereas the y-axis represents the counts of GA+NT.

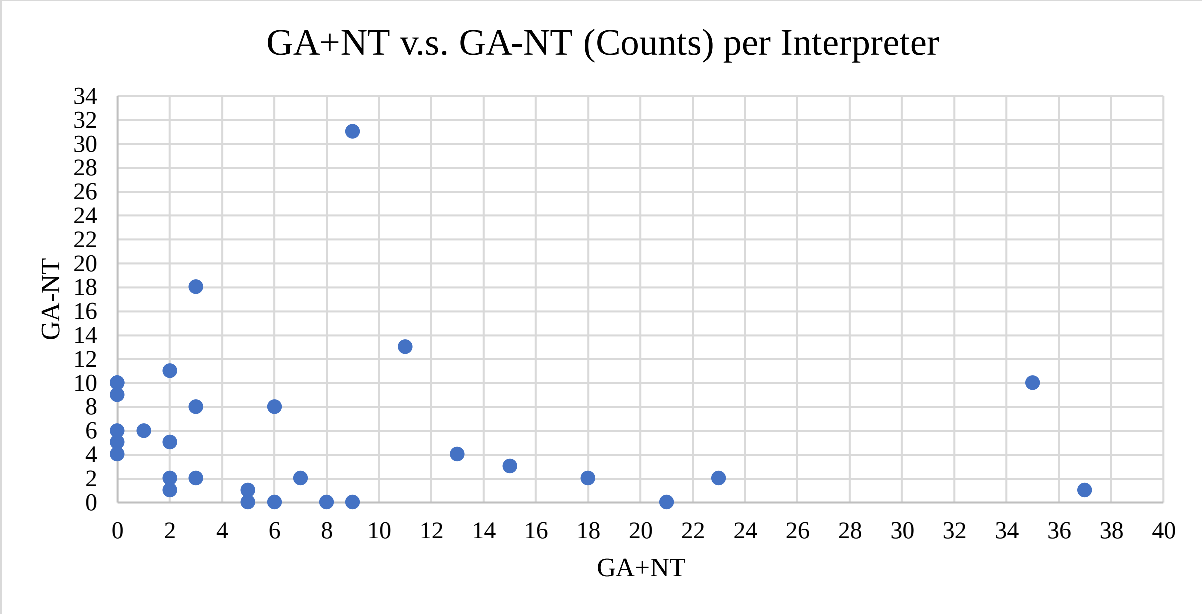


Figure 15 GA+NT vs GA-NT by all interpreters (counts)

A first glance of the chart would reveal that whereas most interpreters performed a mixture of both types of GA, others are associated with only one type of GA, as indicated by the value of zero on either axis. Specifically, 11 out of 30, or 37% of the interpreters, only showed one type of GA. Among these 11 interpreters, 5 did not stop writing when averting their gaze. By comparison, 6 of the 11 interpreters completely detached themselves from taking notes during GA. For the 19 interpreters who were associated with both GA types, there were noticeable individual differences regarding the distribution and preference of each type. Two pie charts (Figure 16) were plotted to demonstrate the distribution of each GA type regarding counts and length in a broad sense to tap into the mixture of each GA type. Overall, the charts evidently suggest that GA+NT was more prevalent among interpreters, with both counts (246 vs 174) and length (302.53s vs 179.52) noticeably exceeding GA-NT.

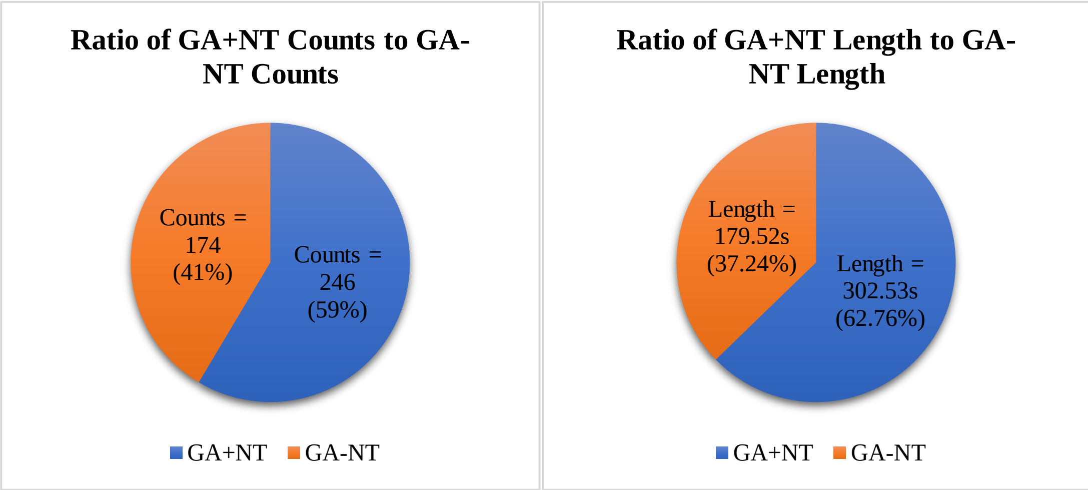


Figure 16 Ratio of each GA type regarding counts and lengths

However, a zoomed-in review of the data revealed variances among the groups. Figure 17 below represents the distribution of each type of GA in terms of total counts and length among interpreters who demonstrated a mixture of each GA type. In terms of counts, more GA-NT than GA+NT was coded from the videos obtained from 2013 and 2014, meaning that interpreters from these two years more frequently averted their gaze from notepads while detaching themselves from handwriting. By comparison, the GA of interpreters from 2015 and 2016 was more frequently accompanied by the action of writing than listening only. Regarding the total length of each GA type, GA+NT in 2013, 2015, and 2016 accounted for the majority of the total time spent in GA, whereas more time was spent in GA-NT in 2014. It is worth noticing that, in 2013, although GA+NT was outnumbered by GA-NT (45% vs 55%), the former cost a longer time than the latter (60% vs 40%). It is also evident from the chart that the proportion of each GA type varied drastically by both measurements.

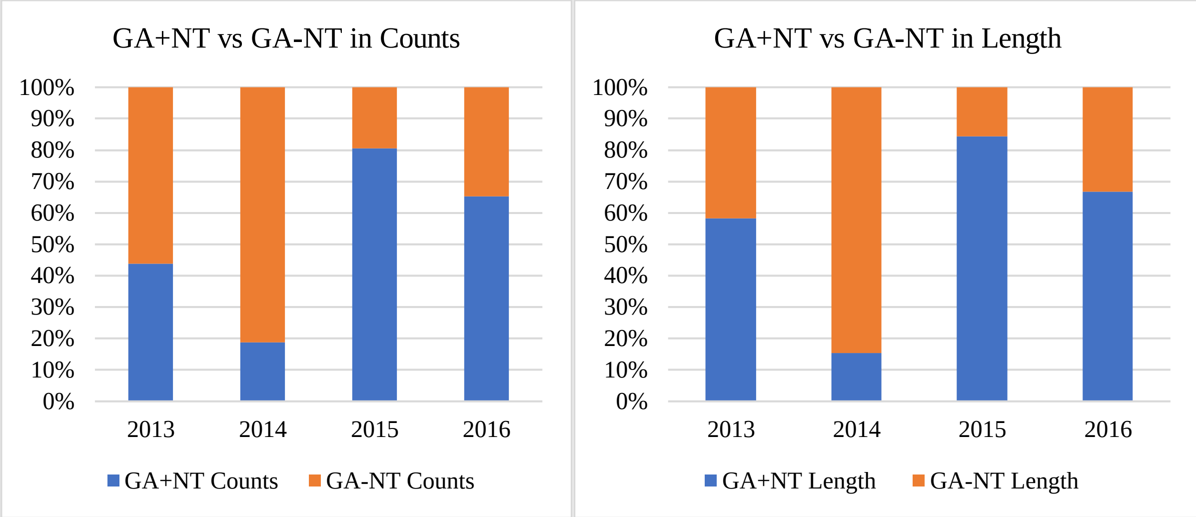


Figure 17 Counts and lengths for each GA type in each year

3.2.1.3.1 Inferential Analysis

Normality Check

For RQ2, the quantitative evidence so far has suggested that GA+NT is more favoured by the participants who demonstrated both types of GA: interpreters collectively spent a much longer time conducting more GA while engaging in handwriting. However, inferential analyses are necessary to test if the discrepancy in counts and length reached statistical significance. A Shapiro-Wilk test was conducted to test the normality assumption for both measurements. In terms of counts, the normality test failed to verify the normality assumption for either GA+NT ($W(19) = .79$, $p = .001$) or GA-NT ($W(19) = .76$, $p = .001$), suggesting that the data should be analysed via a non-parametric test. Given the specific issue to explore and the numbers of the data groups, a Wilcoxon test was chosen to test if there exists statistical significance between the counts of GA+NT and GA-NT. For the measurement of length, prior to the normality check, the mean length of each GA type for each interpreter was computed. The test did not show evidence for non-normality for either GA+NT ($W(19) = .97$, $p = .78$) or GA – NT ($W(19) = .95$, $p = .45$). Based on the outcome and a visual examination of the histogram and Q-Q plot, a paired sample t-test was deemed methodologically sound for the measurement of length.

Wilcoxon Test

A Wilcoxon test did not render evidence for significance between GA+NT ($MD = 8.44$, $n = 19$) and GA-NT ($MD = 10.35$, $n = 19$) in terms of the gross counts, $z = -.79$, $p = .44$.

Paired Sample t-test

Similarly, regarding the mean length, the *t*-test also failed to render statistical significance between GA+NT (M = 1.05, SD = 0.56) and GA-NT (M = 1.00, SD = 0.53), $t(18) = 0.28$, p = 0.78. Combined with the results from quantitative analysis, although quantitative analysis indicated an overall preference for GA+NT over GA-NT among interpreters, qualitative analyses did not identify statistical significance regarding the discrepancies between the two types of GA.

In summary, the data coded for addressing RQ2 suggested that the majority trainee interpreters showed a mixture of GA types, meaning that during GA, they would either engage in or detach from the action of handwriting. But inferential analysis did not find significant effect for any preference.

3.2.2 Textual features preceding GA (RQ3)

3.2.2.1 Overall description

The third RQ investigated in this corpus study concerns the textual features immediately preceding each GA. Specifically, the question is laid out to explore potential GA inducers in the context of CI. It is hypothesised that features which could potentially increase cognitive load would appear in the source texts immediately preceding GA, and, among the textual features, sentences with subordinate clauses (complex sentence structures) and LFWs would be the most popular features.

Previous studies suggested that GA tends to occur in response to high cognitive load (Glenberg *et al.*, 1998; Doherty-Sneddon *et al.*, 2002; Doherty-Sneddon and Phelps, 2005). In interpreting studies, although the topic has not yet been adequately researched, some challenging textual features, such as sentences with embedded clauses, numbers, rare words, and acronyms, have been reported to be associated with an increased demand of processing capacity (Gile, 2009). In this corpus study, to narrow down the potential elements that could lead to GA, each GA was examined separately regarding if the source text segments immediately preceding GA would contain features that could potentially impose a higher cognitive load. Several features were observed, including complex sentence structure, LFWs, negation, number, question, passive voice, and acronyms. The frequency of each textual feature is summaries in Figure 3.5 in 3.2.2.2.

An initial observation of GA and its preceding segments revealed two overall findings. First, a proportion of the source text segments preceding GA did not contain features that would likely require an increased volume of memory resources. For example:

(1) *The idea behind the fiscal cliff was that if the federal government* [GA] *allowed the two plans to go ahead, they would have a detrimental effect on an already shaky* [GA] *economy.* (Extracted from 2013)
(2) *Dismantling outrageous pseudoscientific* [GA] *claims is an excellent way* [GA] *to learn the basics of science.* (Extracted from 2014)

In example (1), while it is relatively clear that the first GA occurred after a predicative clause (introduced by "that") and an adverbial clause of condition (introduced by "if"), the source speech segment before the second GA does not include noticeable features that would be exceptionally demanding. Similarly, in example (2), the first GA occurred immediately after the LFW "*pseudoscientific*" (Zipf value = 1.17) was uttered, whereas the second GA followed a simple and easy text segment that is free from elements that would typically require additional cognitive resources.

Altogether, as shown in Figure 18 below, among the 420 times of GA, 316 times (75.24%) occurred immediately after potential triggers, leaving a quarter of GA (104 times (24.76%)) cannot be linked to any immediate textual features that would typically impose extra cognitive load. An interesting aspect of this finding is that whereas the majority (24 out of 30) of participants occasionally averted gaze when the preceding text did not contain potential triggers, 6 interpreters only conducted GA after experiencing potential GA triggers.

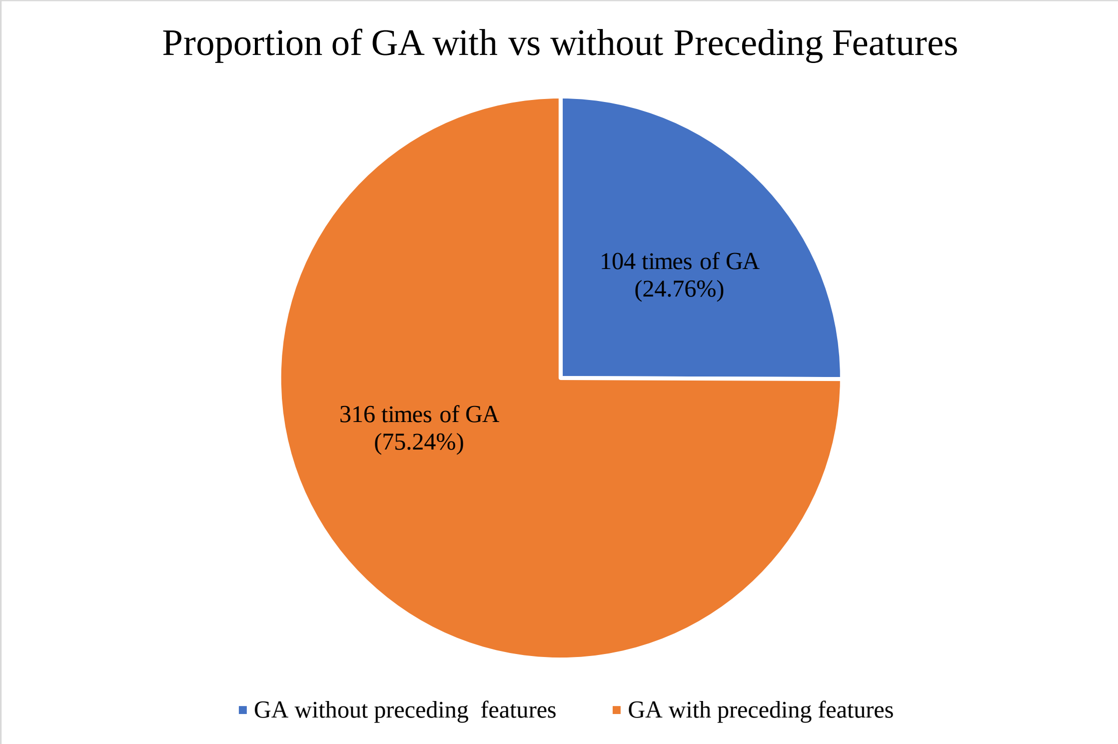


Figure 18 Proportion of GA with and without preceding features

Another overall finding is that for the amount of GA that occurred after the presence of potential triggering textual features mentioned above, the majority of GA was followed after only one of the features was processed, whereas the rest of GA was associated with a combination of the several features. The proportion of the two scenarios is illustrated in Figure 19 below. Most of the GA (69.94%) occurred after only one of the textual features was processed. By contrast, a much smaller portion of GA (30.06%) was preceded by a combination of textual features that could potentially increase cognitive load. The difference sustains even when the general data is zoomed in to the cohort-based level. As shown in the bar chart, across the cohorts, the amount of GA occurring after one of the textual features surpassed its counterpart, accounting for 55% to 75% of the total amount of GA.

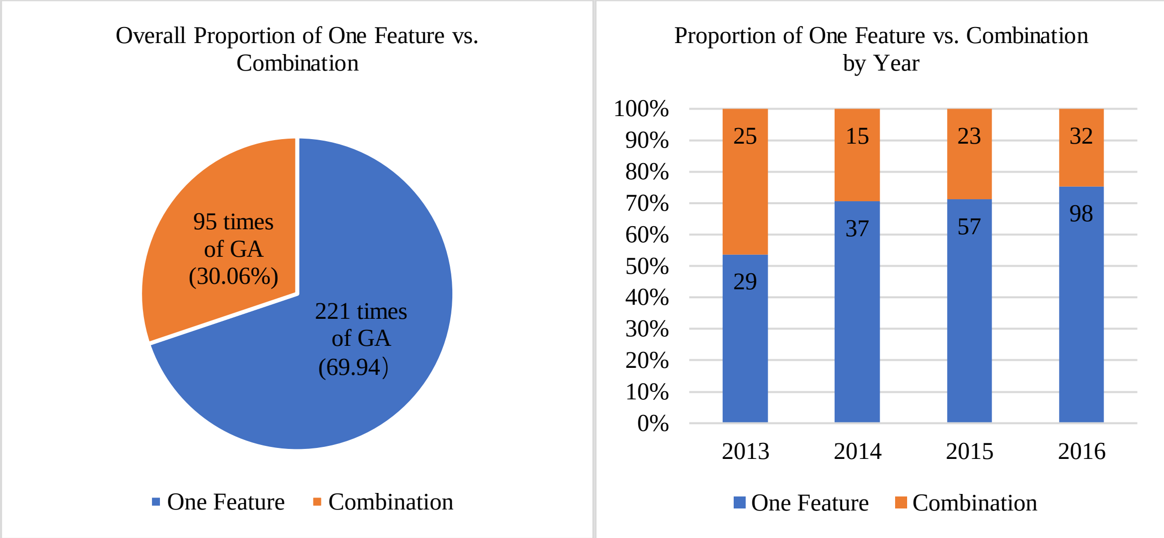


Figure 19 Proportion of one feature vs combination

3.2.2.2 Frequency of textual features

As mentioned earlier, among GA that occurred after the presence of textual features that could potentially cause cognitive load to increase, 220 were preceded by one of the seven features. Table 3.14 below compares the frequency of each textual feature observed from the four source speech. The features are listed in a frequency-descending manner in the table, where the most frequent feature sits at the top of the chart. Complex Sentence Structure dominates the chart with a frequency of 118 times, taking up a proportion of 53.39%. In other words, more than half of the GA in this category can be traced back to a sentence containing subordinate clause(s). The second most frequent feature was LFW which appeared 45 times (20.36%) in total. The rest five features were observed much less frequently, with a proportion ranging between 3.62% and 8.60%.

As previously mentioned, an aspect of the hypothesis is angled toward an association between GA and two textual features – Complex Sentence Structure and LFW. Concerning this aspect, a striking feature from the table is that the two features were the only ones that reached two-digit proportions, and the two features were associated with 70% of GA preceded by a single textual feature.

Table 12 *Frequency and proportions of textual features that appeared alone*

Number	***Textual Features***	***Frequency***	***Proportion***
1	Complex Sentence Structure	118	53.39%
2	LFW	45	20.36%
3	Number	19	8.60%

4	Negation	11	4.98%
5	Question	10	4.52%
6	Passive Voice	9	4.07%
7	Acronym	8	3.62%

In terms of GA that occurred after the presence of a group of features, the seven observed features resulted in 19 combinations summarised in Table 12 in a frequency-descending manner. Perhaps the first noteworthy characteristic observed from the table is the considerable discrepancy in frequency and proportion among the combinations. Whereas some combinations are only associated with a single occurrence of GA, other combinations appeared much more often, resulting in a much higher proportion. Specifically, it is evident that the combination of Complex Sentence Structure and LFW was the most frequent, with nearly a quarter (24.21%) of GA in the category occurring immediately after the combination was present in the source speech. By contrast, the Number + Passive and another 8 combinations were the least frequent, with each appearing once and taking up 1.05%, respectively. The second most frequent combination was composed of Complex Sentence Structure and Question (16.84%), followed by Complex Sentence Structure and Negation (13.68%) and Complex Sentence Structure and Passive Voice (11.58%).

A prominent feature of the table is that Complex Sentence Structure and LFW are elements of several other combinations (number 9 – 13 in the table 8 below). Altogether, 6 out of 19 combinations are composed of the two features, contributing to a proportion of 31.91%. In other words, Complex Sentence Structure and LFW are associated with nearly a third of GA that occurred after a group of potential GA triggers.

Table 8 Frequency and proportion of textual features that appeared in combination

Number	***Combination***	***Frequency***	***Proportion***
1	Complex Structure + LFW	23	24.21%
2	Complex Structure + Question	16	16.84%
3	Complex Structure + Negation	14	14.74%
4	Complex Structure + Passive Voice	11	11.58%
5	Complex Structure + Number	7	7.37%
6	Complex Structure + Negation + Acronym	5	5.26%
7	LFW + Number	4	4.21%

8	LFW + Acronym	3	3.16%
9	Complex Structure + LFW + Acronym	2	2.11%
10	Complex Structure + LFW + Negation	2	2.11%
11	Complex Structure + LFW + Number	1	1.05%
12	Complex Structure + LFW + Negation + Passive Voice	1	1.05%
13	Complex Structure + LFW + Passive Voice	1	1.05%
16	Complex Structure + Number + Question	1	1.05%
14	Complex Structure + Acronym	1	1.05%
15	Negation + Acronym	1	1.05%
17	Passive Voice + Number	1	1.05%
18	Number + Acronym	1	1.05%

Consistent with previously reported findings in this chapter, the textual features observed in table 8 have shown to be associated with considerable individual differences. Specifically, the textual features experienced prior to GA varied drastically among interpreters. Figure 20 below is plotted to compare the number and proportion of interpreters associated with a certain textual feature. A striking detail from the chart is that all interpreters processed complex sentence structures and rare words before averting gaze from notepads. By contrast, other features were only associated with some interpreters. For example, only a third of interpreters averted gaze from notepad after hearing a question. The presence of numbers, a classic element that would burden WM and hinder comprehension (Gile, 2009; Wang, 2015), was only associated with 16 interpreters (53.33%). The proportions of interpreters who averted gaze after negation, passive voice, and acronyms were 56.67% (17), 55.33% (16), and 50% (15), respectively.

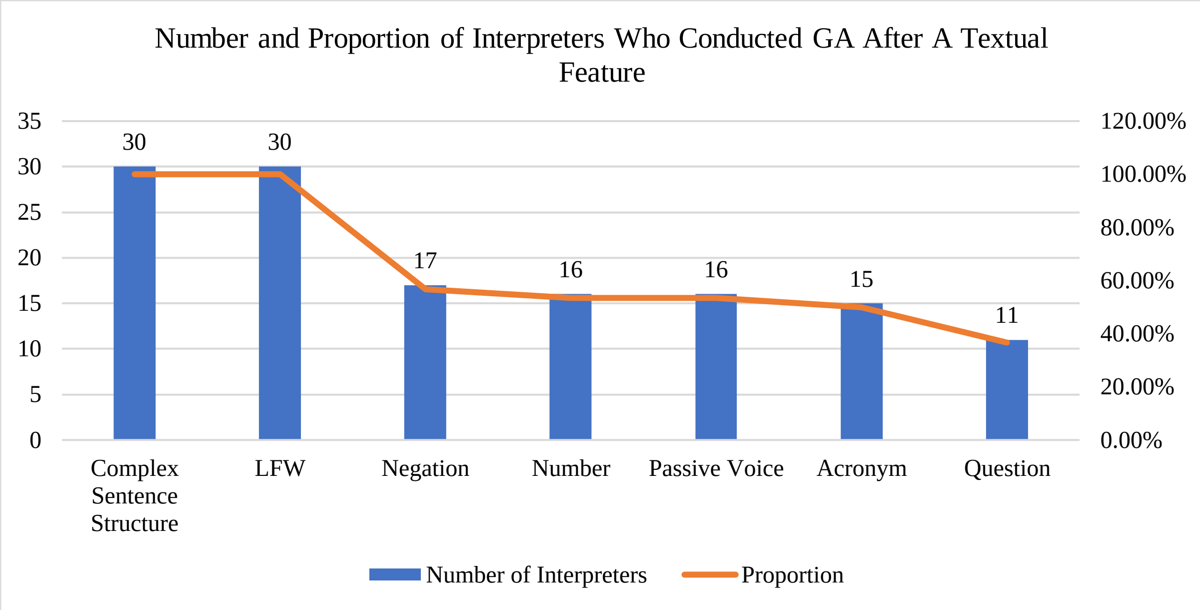


Figure 20 Number and proportion of interpreters regarding each textual feature

So far, descriptive data regarding the frequency and proportion of features, either by single appearance or combination, and the data pertaining to interpreters' individual differences in being affected by the features indicates that Complex Sentence Structures and LFWs could be strong GA triggers. However, as suggested by scholars, quantitative data alone would not suffice to reveal the relationship between the variables (Tashakkori and Teddlie 2003; Creswell, 2005; Ivankova *et al*., 2013. Following the mixed-methods sequential design, quantitative data was generated and analysed. The results are reported in the following section.

3.2.3 Sentence difficulty and GA (RQ3)

3.2.3.1 Overall Description

To test the hypothesis, a scoring system was introduced for GA and non-GA sentences. To recap, each LFW and subordinate clause, each denotes a score of 1. Once all sentences of the four source speech was scored, the mean value of GA sentences and non-GA sentences were computed for each interpreter. Altogether, there are 30 pairs of scores divided in two groups. A box plot (Figure 21) below illustrates the distribution of all mean scores.

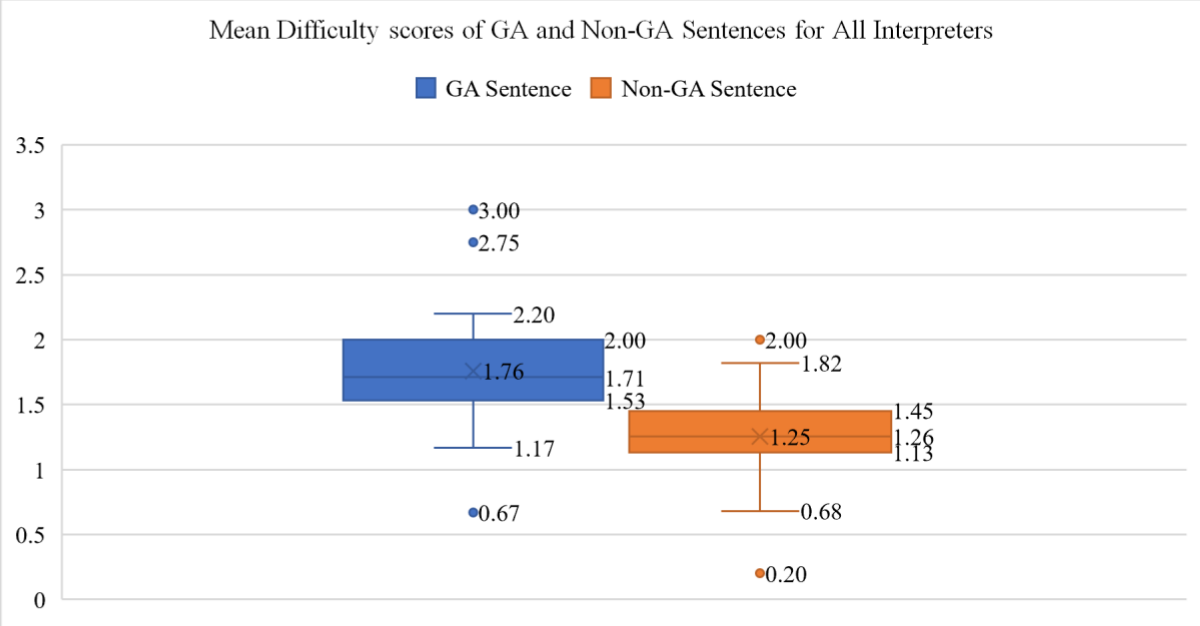


Figure 21 Distribution of mean difficulty scores of GA and non-GA sentences

As indicated in the box plot, the median score for difficulty of GA sentences is 1.71, with an interquartile range between 1.53 (Q1) and 2 (Q3), suggesting that half of the sentences where GA was coded have difficulty scored between 1.53 and 2. On average, the difficulty score of GA sentences is 1.76 (*max* = 3, *min* = 0.67). By comparison, the median score for difficulty of non-GA sentences is 1.26. The interquartile range is from 1.13 (Q1) and 1.45 (Q3), a smaller range than that of GA sentences. On average, non-GA sentences have a smaller score of 1.25 (*max* = 2, *min* = 0.2). Both data categories have outliers on either whisker. It would seem that GA sentences are more difficult than GA sentences, but inferential analyses are required to test if the difference reached statistical significance.

3.2.3.2 Inferential Analysis

3.2.3.2.1 Normality Check

A paired sample *t*-test was conducted to test if GA sentences are significantly more difficult than non-GA sentences. Prior to the analysis, a Shapiro-Wilk (see Table 3.17 below) test was performed to test the assumption of normality. The normality check did not suggest evidence against normal distribution for either GA sentences ($W = .93$, $p = .06$) and non-GA sentences ($W = .96$, $p = .26$). Based on the outcome and a visual examination of the histograms and q-q plots, and considering the number of data groups, adopting a paired sample *t*-test was deemed methodologically reasonable.

3.2.3.2.2 *t*-test

The paired sample *t*-test revealed that GA sentences ($M = 1.76$, $SD = .44$) are more difficult than non-GA sentences ($M = 1.25$, $SD = .36$). The difference, 0.51, has reached statistical significance, $t(29) = 4.52$, $p < .001$. In the context of this corpus study, the result indicates that interpreters are more likely to conduct GA when processing source texts containing either subordinate clauses or LFWs or both.

3.2.4 Summary of Key Findings

First, trainee interpreters averted their gaze from note pads during CI active listening and notetaking, although the total times spent doing so was relatively low. However, there were vast individual differences regarding the frequency and the time spent in GA. Second, inferential analysis did not find any significant effect between GA+NT and GA-NT. Third, GA was associated with various textual features immediately preceding each GA. Complex sentence structures and LFWs were the most common features after which GA was observed – all interpreters performed a certain amount of GA, ~~if not all~~, after experiencing either or both of the features; the two features, either by single appearance or combined, were associated with the most significant amount of GA across the corpus. Inferential analysis suggested that the sentence difficulty, defined by complex sentence structure or LFW or both, was significantly higher for GA sentences than non-GA sentences, indicating that interpreters would be more likely to avert their gaze when processing complex sentence structures and/or rare words.

3.3 Discussion

3.3.1 Overview

So far, the investigation within this initial corpus study has been geared towards answering three RQs. To recap, the first RQ taps into GA in the context of CI active listening and notetaking by asking if interpreters would avert their gaze from notepads at all. RQ2 is laid out to explore if interpreters would remain engaged in notetaking or disengage from notetaking during GA. RQ3 probes into the possible GA inducers by examining the textual features that immediately precede GA. Evidence was gathered and analysed to provide answers to the above questions, and the findings of each RQ are:

1. Answering RQ1 – As expected, trainee interpreters averted their gaze from notepads, but the exact manner of GA regarding frequency and time elapsed in GA varied dramatically due to considerable individual differences.

2. Answering RQ2 – For most interpreters, GA occurred in both scenarios: sometimes interpreters would keep writing notes while looking away from notepads, and other times they would detach from recording notes when GA took place. Inferential analyses suggested that the frequency and time spent in GA for both types were not significantly different.
3. Answering RQ3 – GA was not necessarily associated with one specific preceding textual feature, although most GA took place after one or a few textual features that could potentially increase cognitive load were encountered. The observed textual features included complex sentence structure (sentences with subordinate clause(s)), LFWs, negation, number, question, passive voice, and acronyms. Complex sentence structures and LFWs were the only textual features that all interpreters experienced before performing varying proportions of GA. Inferential analyses revealed that interpreters are significantly more likely to conduct GA when processing sentences containing complex structures and/or LFWs.

Some findings from the study generated overarching significance in understanding GA in CI active listening and notetaking. A General Discussion chapter (Chapter 5) is dedicated to interpreting the overarching findings from this thesis. To avoid repetition, the following section is dedicated to interpreting results specifically relevant to RQ2 and RQ3.

3.3.2 GA with Notetaking vs GA without Notetaking (RQ2)

3.3.2 1 GA with Notetaking

As previously reported in 3.2.2.3, data regarding the action accompanying GA has suggested that some interpreters continued the action of notetaking when looking away from the notepads, meaning that interpreters did not monitor their handwriting while recording information. Studies have suggested that the action of writing typically relies on eye-hand coordination (Kaiser, Albaret, and Doudin, 2009; Ujbányi, Kővári, Aziládi, and Katona, 2020). According to Ujbányi *et al.*, eye-hand coordination enjoys fundamental benefits to handwriting – during writing, the eyes would feed information regarding hand position to the brain, which would subsequently utilise the information to instruct the hand to move in a way that leads to the creation of lines, shapes, and letters. The authors also suggested that by collecting visual information, the brain would detect when errors occur. The absence of visual contact in writing was explored in an MA project (Guo, 2016), where trainee interpreters were instructed to refrain from monitoring handwriting during CI active listening and notetaking. Guo’s (2016) study aimed to investigate if unmonitored CI notetaking (blind notetaking) would impact performance. Trainee interpreters reported that blind notetaking engendered an insecure and unnatural feeling. The most frequently mentioned fear, according to the participants, was that taking notes without monitoring would easily lead to overlapped strokes, which would hinder note reading and, ultimately, the faithfulness of target speech. Based on the above literature and considering the significance of orderly CI notes (Rozan, 2002; Albl-Mikasa, 2017; Gillies, 2019), it would be warranted to argue that averting gaze from notepads while sustaining the action of writing would be unusual to a degree.

However, this seemingly strange behaviour would make sense from two perspectives. First, CI notetaking engenders significant bearings on the faithfulness of interpreting as it entails a deep and comprehensive mental analysis of the source speech. As mentioned in 2.6.2.2, faithfulness is the nucleus of assessing the quality of interpreting (Macías, 2006; Bartłomiejczyk, 2007; Hale, 2007; Setton and Motta; 2007; Ke and Zhang, 2008; Lee, 2008; Gile, 2009; Choi, 2013; Wu, 2013; Pöchhacker 2002, 2015), and it is usually embodied in completeness and accuracy (Hale, 2007; Pöchhacker, 2015).

It would be well-grounded to argue that notetaking benefits completeness. As discussed in Chapter 2, STM/WM systems are limited in capacity and cannot retain the amount of information from the source speech. In other words, solely relying on memory would result in significant information losses. Due to notetaking's storage function, interpreters' notepads would serve as extra memory where essential information is stored (Albl-Mikasa, 2016; Pöchhacker, 2016). As suggested by scholars (e.g., DiVesta and Gary, 1972; Kiewra, 1985; Poverly and Wolf, 2019), the storage function of notes facilitates recall via note reviewing. According to Kiewra (1985), while notetaking cannot guarantee successful recall, the absence of notes could almost result in an inability to recall. In the context of CI, the efforts invested in taking notes would be rewarded with the opportunity to use notes as memory cues (Setton, 2002), ultimately contributing to a satisfactory completeness.

Taking notes would also improve accuracy. As previously discussed, another function of notetaking is speech encoding (DiVesta and Gary, 1972; Kiewra, 1985). According to Kiewra, by encoding, the notetaker processes the speech while "transcribing whatever subjective associations, inferences, and interpretations occurred to him while listening" (p. 8). From the perspective of CI, the encoding process would facilitate the construction of sense, an essential aspect of accuracy where interpreters consciously render the linguistic meanings of the words while processing the non-linguistic aspects of the speech (Seleskovitch, 1978).

It would be warranted to argue that faithfulness was of extra significance to the trainee interpreters in the videos coded in this corpus study. To recap, the videos recorded CI final examinations where the trainee interpreters' performance was assessed. At Newcastle University, the institute where the trainee interpreters were registered, faithfulness carries 70% of the weight in the marking criteria. To avoid significant loss or distortion of information, it would seem risky for the trainee interpreters to deviate from the action of taking notes.

The second possible reason to explain why trainee interpreters remained engaged in notetaking even when the eye-hand coordination was unavailable lies in the perspective of cognitive load levels. As mentioned in 2.2.3.3, cognitive load is not static but dynamic (Whelan, 2007; Brünken *et al.*, 2010). In the context of interpreting, according to the Effort Model (Gile, 2009), an upsurge in cognitive load could lead to performance deterioration. Gile suggested that during the first phase of CI, WM resources are concurrently consumed in four activities: listening and analysis, notetaking, STM operations, and coordination. The four components compete for the limited supply of WM resources and could lead to performance deterioration when cognitive overload occurs, a situation where interpreters are advised to prioritise active listening. According to the Cognitive Load Hypothesis (Glenberg *et al.*, 1998), people would look away from distractions to modulate high cognitive load. Looking away is beneficial as it would lead to a deduction of information that needs processing and free up some WM resources to be directed to the task at hand. In the context of the corpus study, when trainee interpreters averted their gaze from notepads, they likely experienced high cognitive load. However, by looking away from notepads, they brought down the level of cognitive load to the degree that the CE could still permit notetaking to continue. Put simply, trainee interpreters continued to record notes during GA because it is cognitively attainable.

3.3.2.2 GA without Notetaking

In previous sections, it has been highlighted that notetaking is vital to CI performance, and interpreters would endeavour to take notes even though they cannot monitor their handwriting. However, some trainee interpreters in this corpus study were also found to detach themselves from the action of handwriting during GA. Although the total counts of GA-NT were smaller than GA+NT, its appearance would immediately raise the question that if notetaking is paramount to the faithfulness of the delivery and if GA would free up enough cognitive resources to maintain notetaking, why would interpreters still detach from recording information in notes? Indeed, GA without notetaking would seem contradictory to the previous reasoning. However, the action can be explained from two angles.

The first possible reason is that pauses from notetaking were deemed necessary at some points. As discussed in 2.6.3.3, the notetaking process in CI is akin to writing in terms of composition and at levels of words, sentences, and discourse (Kirchhoff, 1979; Kohn and Albl-Mikasa, 2002; Chen, 2016; Albl-Mikasa, 2017, 2019). Studies on writing have indicated that pauses can take up almost half of the writing time and usually happen when higher-level cognitive processes occur (Alamargot, Dansac, Chesnet, and Fayol, 2007; Alves, Castro, Sousa, and Strömqvist, 2007; Olive, Alves, and Castro, 2009). According to the scholars, pauses during writing could signal planning, revising, retrieving, and monitoring (Scholperoord, 2002; Olive *et al*., 2009). Within this corpus study, it would be warranted to argue that trainee interpreters did not monitor or try to revise their notes during the pauses as they were not looking at the notepads. An alternative and more feasible explanation would be that the interpreters paused were trying to retrieve the meaning of the word or come up with a conversion plan. See the following examples:

Example 1

> *Their knowledge of science is* ***rudimentary***. (Extracted from the source speech used in 2014).

Moments after "*rudimentary*" was uttered, 80% of the interpreters in 2014 paused from notetaking while averting their gaze from notepads. It can be immediately noticed that the sentence is short and constructed in a simple structure, which, according to Liu and Chiu (2009) should be easy to process. A possible explanation for the note-less GA could be that the word "*rudimentary*" made notetaking difficult. According to the SUBTLEX-UK database (Hueven *et al*., 2014), "*rudimentary*" has a Zipf value of 2.92, making it a word with low frequency.

Syntactically, the word shoulders the role of the predicative adjective of the sentence, and therefore, knowing and remembering the meaning of the word would be vital for this short sentence to be faithfully rendered. However, considering the WFE, the word, if known to the interpreters, would require a relatively long time to convert to note. The rarity of the word would contribute to the possibility that the interpreters had not developed a symbol or abbreviation for the word, meaning that they would need time to conjure up a plan for conversion. In this instance, the pause from notetaking could likely result from trying to retrieve the meaning of "*rudimentary*" and coming up with a conversion plan. In other words, interpreters could not engage in the act of taking notes because they did not know what to write.

Another possible explanation for the occurrence of GA-NT could be that the interpreters experienced high cognitive load to the degree that averting their gaze from notepads was simply inadequate to modulate the cognitive load. As discussed above, when GA would suffice to modulate cognitive load to a level where interpreters can manage to engage in handwriting, they would endeavour to do so to ensure better completeness and accuracy. However, if required resources are in deficit even after GA was performed, more resources must be diverted to the task to avoid failure. Out of the four components laid out in the Effort Model, notetaking would arguably be the only candidate from which WM resources can be diverted. The nature of CI requires interpreters to prioritise listening and analysis at all times by investing unmitigated attention (Gile, 2009; Herrero, 2017). Coordination must also be guaranteed, as it is fundamental to managing attentional resources (Leeson, 2005) and balancing the allocation of resources (Kriston, 2012). It would even seem superfluous to highlight the significance of memory operation as mounting evidence has suggested that memory systems underpin cognitive tasks (e.g., Atkinson and Shiffrin, 1968; Baddeley and Hitch, 1974; Baddeley, 2009; Cowan, 2009). Gile (2009) suggested that cognitive resources spent on notetaking should be controlled at a minimal level to prioritise active listening. Similarly, Scholperoord (2001) suggested that cognitive overload is one reason that leads to pauses during writing. See the following example:

Example 2

> *So dismantling outrageous* ***pseudo****-scientific claims is an excellent way to learn the basics of science, partly* ***because*** *science is largely about disproving theories but also* ***because*** *their lack of scientific knowledge among miracle cure therapists and journalists gives us some very simple ideas to test.* (Extracted from the source speech used in 2014)

The majority of interpreters (80%) from 2014 averted their gaze without taking notes during the utterance of the sentence. The lengthy sentence contains one LFW (pseudo, Zipf value = 2.74) and two subordinate clauses, with each introduced by "*because*". Interpreters swiftly conducted GA-NT after "*pseudo*" was uttered. Again, the reason for interpreters to pause notetaking at this particular moment would likely be that they were searching for the word in their mental lexicon and subsequently endeavouring to plan on the conversion. As the source speech continued, interpreters would soon notice that the syntactic structure of the sentence being uttered was not easy.

Based on the mark-ups from the corpus, interpreters averted their gaze from notepads and paused from notetaking during the utterance of the clause subject "*their lack of scientific knowledge among miracle cure therapists and journalists*". It would be warranted to argue that the sentence's subject is lengthy and dense in information, which, according to Liu and Chiu (2009), could indicate high speech difficulty that contributes to a high level of cognitive load. From the perspective of taking notes, the subject could pose a challenge to interpreters. To grasp the SVO structure and attain the principle of verticality (Ilg and Lambert, 1996; Rozan, 2002; Kohn and Albl-Misaka, 2002; Jones, 2014; Gillies, 2017, 2019), considering the incremental process of listening comprehension (Tanenhaus *et al*.'s 1995), the interpreters could not thoroughly render the subject until the verb "*gives*" was heard, meaning that the time pressure for recording this piece of information would be high. Also, since animacy serves as an important processing cue in Mandarin Chinese relative clauses (Wu, Kaiser, and Anderson, 2012), it is likely that interpreters would need to restructure the sentence in the delivery to "*miracle cure therapists and journalists lack scientific knowledge*" so that the target speech would sound natural to the audience. Ideally, this inverted rendition should be reflected in notes. For instance, interpreters could use lines and arrows to indicate the order. Put simply, the time urgency, the necessity of thoroughly analysing the subject, and the need for planning might have jointly exposed some interpreters to a high cognitive load level that not only induced GA but also expropriated the effort from notetaking.

3.3.3 Possible GA Inducers (RQ3)

As reported in 3.2, a range of textual features preceding GA was observed to narrow down possible GA inducers in the context of CI active listening and notetaking. In total, seven features were observed, including complex sentence structure, LFWs, negations, numbers, questions, passive voice, and acronyms. The features either appeared alone or in combinations, and some of the features appeared much more frequently than the others.

All interpreters had a proportion of GA performed after experiencing complex sentence structures and LFWs. By comparison, the rest of the features appeared in the source text before GA for some interpreters. Complex sentence structures and LFWs might outweigh other features in terms of the possibility of inducing GA. Quantitative and inferential evidence has suggested that the two features not only appear more often than the rest but also contribute to higher difficulty for GA sentences. In the following section, the features are discussed regarding their potential to increase cognitive load in the context of CI active listening and notetaking. The order of discussion indicates the high-low frequency of each feature.

3.3.3.1 Complex sentence structure

Complex sentence structures are considered one of the problem triggers during interpreting (Gile, 2009), a notion well reflected in this corpus study. The finding that complex sentence structures could be one of the GA inducers indirectly accords with previous studies where mounting evidence was generated towards the influence of syntactic complexity on sentence comprehension (Just and Carpenter, 1992; Ferreira *et al*., 1996; Gibson, 1998, 2000; Fedorenko *et al*., 2004; Fallon *et al*., 2006). These studies traditionally compared objective relative clauses with subjective relative clauses and reported that the former is more cognitively expensive than the latter because the former would typically require the retention of information in WM until relevant obligatory syntactic components are integrated into the mental representation. In this corpus study, the source speech did not contain any objective relative clauses where the arguments and their obligatory heads are kept at a long distance. Instead, the speech contained sentences that still qualify as syntactically complex but not as challenging as objective relative clauses that were used in the above-mentioned studies. See the following example:

Example 3

> Source speech: *The idea* ***behind the fiscal cliff*** *was* ***that if*** *the federal government allowed these two plans to go ahead, they would have a detrimental effect on an already shaky economy, sending it back into an official recession* ***as*** *it would cut household incomes, increase unemployment rates and undermine consumer and investor confidence.* (Extracted from the 2013 source speech)

The majority of the trainee interpreters in 2013 averted their gaze from notepads during the above sentence was uttered. An initial observation would reveal that the sentence is embedded with four subordinate clauses. The first is an attributive clause "*behind the fiscal cliff*" following and describing "*the idea*". The second is a predicative clause introduced by "*that*". The third one is an adverbial clause of condition introduced by "*if*", and the last one is another adverbial clause of cause introduced by "*as*". Most of the GA occurred during the "*that if*" clause. GA was coded once after "as" and zero after "*behind the fiscal cliff*". One would argue that the sentence, albeit long and embedded with subordinate clauses, would be straightforward to comprehend because, by the look of it, the information would flow smoothly in an easy-to-grasp logic. However, in the context of CI, the sentence could be challenging for interpreters for several reasons.

First, interpreters must comprehend the sentence via active listening while endeavouring to convert the sentence into notes. Cutler and Clifton (1999) suggested that processing spoken language does not enjoy the privilege of seeing punctuation marks or spaces, meaning that, in this case, the interpreters had to rely on syntactic cues to segment the continuum of speech into smaller meaningful units. According to the principle of incrementality (Tanenhaus *et al*. 1995), as the speech continues, interpreters must keep updating the mental representation of the sentence to avoid miscomprehension. From the perspective of WM, according to Baddeley (2000), the mental representation would be kept in EB while linked to LTM. Since EB can only hold up to four episodes (Cowan 2005; Baddeley, 2000), the more information that must be integrated into the mental reorientation, the higher WM load will be.

The above-mentioned cognitive price of sentence processing, according to Gibson (1998, 2000), is dissected into a memory cost and integration cost. According to Gibson, making grammatical sense of sentences requires remembering and integrating necessary syntactic components, and the cost will increase with the number of information that needs to be remembered and integrated. The minimum number of syntactic components that must be held in memory is only two: the head noun and the head verb (predicate). For the sentence in the example, the minimal structure would be "*the idea + that*". However, the information contained in this structure is far from what the sentence is intended to express, meaning that the interpreter must integrate more information and hold the information that was previously added to the mental representation. For instance, to comprehend the subject of the sentence, the interpreter must gradually integrate and hold "*behind the fiscal cliff*" in memory until the utterance of the linking verb "*was*". Similarly, thoroughly comprehending the long and complex predicate of the sentence would propel the interpreter to gradually construct a mental representation where two sets of "*noun + predicate*" structures must be held and complemented. As mentioned in 2.1, intrinsic cognitive load is inherited from the material and cannot be altered. From the perspective of how cognitive load is constructed, processing the sentence in the example would firstly impose a high level of intrinsic cognitive load due to the high interactivity between elements.

According to CLT (Sweller *et al.*, 1988; Sweller and Chandler, 1994; Sweller, 2011; Galy *et al.*, 2012), the level of cognitive load is also affected by the level of extraneous cognitive load that is stemmed in the manner of presentation and the extra activities that one must perform alongside the main task. As previously mentioned, CI interpreters usually perform notetaking while conducting active listening. Therefore, the examination of interpreters' cognitive load will not be thorough without including the impact of notetaking. The above authors suggested that the more effort invested in the side task, the higher the extraneous cognitive load level will be. Since interpreters would endeavour to produce a set of notes that reflect the sense and structure of the source speech, for notetaking to proceed smoothly, a certain amount of effort must be guaranteed, leading to a certain level of extraneous cognitive load which, combined with intrinsic cognitive load, contributes to the gross consumption of cognitive resources.

A critical aspect of notetaking is that the task itself is cognitively demanding, meaning that it is also susceptible to intrinsic and extraneous cognitive load. Although it would be warranted to assume that extraneous cognitive load would be relatively low in CI notetaking since it does not involve side tasks, the intrinsic cognitive load for CI notetaking would be at the mercy of material difficulty. Therefore, the above-mentioned cognitive costs, which are derived from having to keep an incomplete mental representation active in memory while integrating more components into the representation, would be reflected in the difficulty in converting source text into notes. In this example, most GA occurred moments after "*that if*" was uttered. While engaging in mentally processing the clause, the interpreters would struggle to note this sentence in a manner that reflects the semantic and syntactic relation between the adverbial of condition and its corresponding result because they might not have schemata with which the long predicate can be recognised and processed as a whole. According to Marcus *et al*. (1996), the absence of schemata would result in a significant consumption of WM and a high level of cognitive load.

Second, during active listening as listeners mainly rely on phonological cues to segment the speech. Phonological interferences have been found to interfere with listening comprehension, a perspective from which can also explain why a cluster of GA was coded soon after "*that if*". As mentioned in 2.6.2.4, a wealth of studies has demonstrated that phonological interferences can hinder word recognition (e.g., Rubenstein, *et al*., 1971; Pexman *et al*., 2001; Lagrou *et al*., 2011; Hartsuiker *et al*., 2013; Rose *et al*., 2015). In the example sentence, "*that if*" was pronounced almost as a whole word. Scholars have suggested that word recognition during speech comprehension entails a memory search process (Marslen-Wilson and Tyler, 1999; Otake and Cutler, 2013; Grosjean, 2018). Specifically, interpreters would access explicit memory where semantic and linguistic knowledge is stored. Until interpreters realised that "*that if*" is not a word but two clause introducers adjacent to each other, they would have aspired to retrieve this word from semantic memory, which requires WM resources (Cohen and Squire, 1980; Schacter, 1987; Baddeley, 2009).

3.3.3.2 LFWs

The potential of LFWs in contributing to high cognitive load is well reflected in previous studies. As mentioned in 2.6.2.4, words with low frequencies take a longer time and more effort to process and recall.

It should be first established that word recognition, regardless of word frequency, carries a cognitive price. As mentioned in 2.1.2.1, knowledge such as vocabulary, schemata, and word knowledge, is stored in semantic memory, a memory store that belongs to explicit memory (Tulving, 1972; Cohen and Squire, 1980; Schacter, 1987; Baddeley, 2009). It is named explicit memory because retrieving information from this store, according to scholars, requires explicit effort in selecting strategies (Shiffrin and Atkinson, 1969; Graft and Schacter, 1985; Squire, 1992; Baddeley, 2009; Vakil *et al*., 2018; Wang, 2020). When accessing explicit memory, effort is also needed to overcome interferences (Fernandes and Moscovitch, 2000) and unwanted responses (Jacoby *et al*., 1989; Kane and Engle, 2000). In the scenario where a word is to be retrieved from semantic memory, one would inevitably invest some cognitive resource to engage in memory search.

However, the cost entailed by word recognition would be higher when retrieving words of low-frequency, a notion lying at the core of WFE (Forster, 1976, 1981; Brysbaert *et al*., 2018; Vitevitch *et al*., 2018; Neville *et al*., 2019; Popov and Reder, 2020, see 2.6.2.4). As mentioned earlier, mental lexicon rests in semantic memory that belongs to explicit memory.

According to scholars, words in one's mental lexicon are stored in descending order, with the highest frequency word stored at the top and the lowest at the bottom, and that the cognitive load induced from searching for a word is influenced by its exact location (Forster, 1976, 1981) – words stored at the lower level would impose a higher level of cognitive load because the listener must invest more cognitive effort to engage in memory search. In this corpus study, LFWs are another frequently observed feature proceeding GA. See the example below:

Example 4

> Sentence 1: *Two Italian economists compared data on* ***Florentine*** *taxpayers from 1427 against tax data from 201,1 and found about 900 surnames still present in Florence*. (Extracted from the 2016 source speech)
> Sentence 2: *It's clear to us all how austerity has* ***maimed*** *the quality of life in Greece and the UK's economy*. (Extracted from the 2013 source speech)

According to the SUBTLEX-UK database (Hueven *et al*., 2014), the Zipf value for "*Florentine*", "*maimed*", and "*stifled*" are 2.82, 2.7, and 2.64, respectively, meaning that the three words are rarely seen.

In 2016, 8 out of 9 interpreters averted their gaze during or immediately after "*Florentine*" was uttered. For interpreters who averted their gaze during the utterance, they began to do so when the syllable "*tine*" was uttered, suggesting that they started processing the word before the utterance was completed. This finding fits well with the Cohort Model (Marslen-Wilson and Tyler, 1980) reviewed in 2.6.2.3. According to the model, word recognition process would commence upon the onset instead of the completion of an utterance. The process is interactive and competitive in a way that the utterance would activate a repertoire of candidate words. As more syllables are heard, the possibilities would narrow down until a winning word is found. In the example, the fact that the interpreters averted their gaze at "*tine*" would also suggest that they identified the word as unfamiliar before the completion of its utterance. A possible scenario is that the interpreters are familiar with the word "*Florence*", which would suit the context as the word "*Italian*" was uttered at the beginning of the sentence. However, when the syllable "*tine*" appeared instead of "*ce*", the interpreters were alerted that the word they were hearing was not the word they anticipated. Startled by the mismatch, the interpreters would begin a memory search in their mental lexicon. Since it is a rare word, searching for the word while processing the new information from the source speech would severely tax WM, resulting in high cognitive load. To circumvent cognitive overload and channel more resources to the search for the word, interpreters strategically resorted to GA. The interpreters who averted their gaze after the full utterance of the word would arguably have gone through a similar path, but the word recognition process for them could be slower.

Similarly, in the second sentence, 4 interpreters in 2013 averted their gaze when "*maimed*" and "*stifled*" were uttered. The reason behind GA, in this instance, it is very similar to the above-mentioned rationale: searching for the two rare words in their mental lexicon resulted in a high cognitive price that was hyped by the need to process the newer information. However, it is noticeable that the two rare words were presented in a complex sentence where a subjective clause was introduced by a formal subject "*It*". Indeed, the syntactic complexity, albeit arguably mild, would have already escalated interpreters' cognitive load to a certain level before the two LWFs entered the interpreters' cognitive system.

3.3.3.3 Number

Mark-ups from the corpus have shown that a proportion of GA was associated with numbers, a well-known problem triggers according to several scholars (e.g., Braun and Clarici, 1996; Mazza, 2008; Giles, 2009). However, compared with complex sentence structures and LFWs, numbers were observed much less frequently.

Numbers are difficult to tackle for several reasons. First, numbers are unpredictable, meaning that interpreters cannot use anticipation as a strategy to pre-analyse the incoming string as they do to syntactic or semantic information. Instead, interpreters must dedicate a considerable volume of cognitive resources to processing the number (Mazza, 2008). Second, numbers can contain high information density, which typically requires more WM resources to process (Alessandrini, 1990; Gile, 2009). Third, numbers are difficult to bind into context, and therefore interpreters would need to change processing strategies from sense oriented to literal translation (Lederer, 1982; Gile, 2009), suggesting a disruption of mental activity. Additionally, pertaining to English-to-Chinese interpreting, numbers in English and Chinese follow different scales, meaning that interpreting between the two languages sometimes compels interpreters to undergo mathematical conversions, which also requires cognitive resources. See the following example:

Example 5

Sentence 1: *On December* ***31, 2012****, tax cuts and across the board government spending cuts are scheduled to become effective*. (Extracted from the 2014 source speech)

Sentence 2: *I later found out that descendants of Japan's Samurai remain elites* ***140*** *years after the ancestors gave up the swords*. (Extracted from the 2016 source speech)

According to Mazza (2008), dates are a form of number that is slightly easier to process than the ones that indicate value. In the first sentence, three occurrences of GA happened around "*31, 2012*". One of the reasons to explain GA, in this instance, could be that the interpreters' minds were geared to comprehending the speech in a contextual sense, and the sudden appearance of a number compelled them to switch to literal translation. The sudden and unexpected switch, according to Gile (2009), requires additional mental efforts. In the second sentence, some interpreters averted their gaze after the presence of "*140*". Apart from the typical reasons why numbers can be challenging, it would be warranted to assume that the cognitive load of the interpreters was already escalated prior to the utterance of the number. It is noticeable that the number was preceded by an indicator of subordinate clauses "*that*". In other words, the interpreters were already exposed to a complex sentence structure before facing the challenge of a number. Also, although “*Samurai*” is not an LFW (Zipf = 3.08), given its nature as a proper name and cultural specificality that would require one’s common knowledge stored in LTM (Meyer, 2008), it would be warranted to argue that the word also cost some extra resources. Considering that cognitive load is additive (Sweller *et al*., 1988) and that WM is limited in capacity, the number "*140*" could increase the cognitive load to the level where GA was necessary to avoid WM overload.

3.3.3.4 Negation

Negation was also observed as a feature preceding GA. Studies found that negative sentences entail higher processing costs. For example, compared with affirmative sentences, negative sentences are more difficult to verify (Just and Carpenter, 1975; Reichle, Carpenter, and Just, 2000). Also, when processing negative sentences, people would ignore the negation as if the sentence is positive (Dale and Durn, 2011; Hasson and Glucksberg, 2006). However, it seems farfetched to associate GA solely to the negative sentences as it has been suggested that the adverse effects of negation should only sustain when negation is processed without a context

(Glenberg, Robertson, and Jansen, 1999; Orenes, Moxey, Scheepers, and Santamaría, 2016). In other words, when negative sentences are spoken in a context, they should entail similar processing cost, if not the same, as affirmative sentences. It would seem superfluous to highlight that interpreters, for the sake of faithfulness and sense construction, would endeavour render within in the context of the speech. Therefore, it would be warranted to argue that the small number of GA related to negation might have resulted from other elements that could increase cognitive load. See the following example:

Example 6

> Sentence 1: *So my dream is to take this issue, **not** just from the compassion argument, to the finance ministers of the world, and say we **cannot** afford to **not** invest in the access to adequate, affordable nutrition for all of humanity.* (Extracted from the 2015 source speech)
> Sentence 2: *I read an article in the Wall Street Journal today and I **can't** think of a better piece of news to begin my talk with.* (Extracted from the 2016 source speech)

The sentences above were associated with small numbers of GA. Specifically, GA was coded three times after the first "*not*" in sentence 1 and twice after "*can't*" in sentence 2. It is noticeable that the first sentence contains three negations, and it would arguably be the case where the latter two negations were more difficult because cause double negations entail higher processing costs than single negations (de-Dios-Flores, 2019). However, GA was coded after the easier single negation instead of the more difficult double negation. A possible explanation is that "*not just from the compassion argument*", as a parenthesis, interrupted the flow of information, which startled the interpreter. To verify the interruption, the interpreter would need to reanalyse the syntactic structure, which increases the level of cognitive load. Sentence 2 was composed of two syntactically simple sentences joined by "*and*". There is a possibility that GA happened due to the high processing costs associated with negation, but the reason does not accord with the finding that negation doesn't entail higher costs when it is processed in a context (Glenberg *et al*., 1999; Orenes *et al*., 2016). "*Wall Street Journal*", however, might have contributed to the increase in cognitive load. Meyer (2008) suggested that processing proper names relies on someone's common knowledge. To access the knowledge, interpreters would need to access their LTM. If interpreters are not familiar with the name, then more effort would be inevitably required (Gile, 2009).

3.3.3.5 Question

A small number of GA was associated with questions. The fact that GA occurred after questions were articulated is not surprising because GA was originally studied in question-answer situations (Glenberg *et al*., 1998; Doherty-Sneddon *et al*., 2002, see 2.3.2.2). However, since interpreters were not there to answer questions but to render the question in the target language, GA, in this situation, was arguably not a result of aspiring to find an answer. See the example below:

Example 7:

> Sentence 1*: Am I middle-class because I start to replace the word 'dinner' with the word 'tea'?* (Extracted from the source speech used in 2016)
> Sentence 2: *Does it make any difference to your class calculator when you hear 'How d'you do?' versus 'Pleased to meet you'?* (Extracted from the source speech used in 2016)

In sentence 1, a cluster of GA was coded during or after the utterance of "*dinner*" and "*tea*". A possible explanation would be that the interpreters were bombarded with a word "*tea*" and "*dinner*" that does not immediately fit in the context of the sentence. To make sense of the speech, the interpreters would need to first process "*dinner*" and "*tea*" literally before binding them into the context. This switch of processing strategy, as mentioned in 3.3.3.3, would impose cognitive load as it relies on CE (Lederer, 1982; Gile, 2009). In the second sentence, instead of switching the processing strategy toward processing a word, the interpreters had to change the trajectory of comprehension towards processing two questions embedded in the sentence. As indicated in 3.3.3.1, The ease of processing a sentence is influenced by the effort invested in building and holding a mental representation in memory. Therefore, it would be warranted to assume that having to fit an additional syntactic structure of a sentence into a mental representation would severely tax, if not overload the WM.

3.3.3.6 Passive Voice

Passive voice was another textual feature that was associated with a fraction of GA. Some studies reported that passive voice entails higher cognitive costs (Gough, 1966; Mehler, 1963; Ferreira, 2003; Millar, Budgell, and Fuller, 2013; Mustafíc, 2020). Ferreira suggested that passives are easily misinterpreted, especially when they describe implausible events. According to Mustafíc, compared with active voice, sentences composed in passive voice tend to signal less precise and clear meanings, implying that processing passives entail higher cognitive costs than processing active voice. See the flowing examples:

Example 8

> Sentence 1: *Much remains to be done, and much more remains debatable.* (Extracted from the 2013 source speech)
> Sentence 2: *It is firmly believed that education is the closest thing we have to a silver bullet when it comes to social immobility*. (Extracted from the 2016 source speech)

A fraction of GA was coded shortly after the two passive voice structures were present. Despite the exiguity, its occurrence was not expected, not only because it is ubiquitous in daily communication but also because the interpreters in the study were MA level students who would arguably be accustomed to processing and expressing ideas in passives. Perhaps one way to examine the influence of passive voice in the above sentences is to reconstruct the sentences in an active manner. Sentence 1, therefore, would be "*We need to do much, and we need to debate much more*", and sentence 2 would be "*We firmly believe that education is the closest thing we have to a silver bullet when it comes to social immobility*". The active sentences are arguably more straightforward to process in terms of comprehension and notetaking. As previously mentioned in 2.6.3.4, a critical aspect of CI notetaking is that notes reflect the comprehensive mental analysis of the speech (Seleskovitch, 1978; Albl-Mikasa, 2008; Setton, 2003; Someya, 2017). Also, interpreters are encouraged to take notes in a preferable vertical and diagonal manner where the basic relations among sentence components are arranged in the SVO order to facilitate smooth delivery of the target speech (Rozan, 2002; Albl-Mikasa, 2017; Gillies, 2019). An arguably easier situation where interpreters can attain the SVO structure with ease would be the three components that enter interpreters' cognitive radar in that order. In the active sentences, since it is explicitly shown that the subjects control the verbs that are imposed on the objects, interpreters would be able to quickly deduce the syntactic and semantic relations and reflect them in notes. By contrast, in the passive voice sentences, the subjects are, in fact, the bearers of the verbs, meaning that a detour would be necessary to attain the active voice formulation both in mental representation and notes. For instance, in Sentence 1, “Much more” is the subject, but it is also the destination of the verb “done [do]”. Although the detour might only enlist an insignificant volume of cognitive resources, it could cause cognitive overload during a cognitively complex activity (Gile, 2009).

3.3.3.7 Acronyms

Acronyms were the least frequent textual features observed in this corpus study. In the sentence in the following example, each acronym was tailed with GA.

Example 9

> Sentence 1: *According to **OECD** figures, Britain may have some of the lowest levels of social mobility.* (Extracted from the 2016 source speech)
> Sentence 2: *I have had the unique and life-changing opportunity to travel with the **UN** World Food Programme and **UNICEF** to visit countries that are affected by poverty and hunger.* (Extracted from the 2015 source speech)

According to Gile (2009), acronyms might not be cognitively costly to process but are more susceptible to insufficient availability of processing capacity. However, for the trainee interpreters whose GA was coded in this study, the processing cost associated with the acronyms might be higher than what Gile has proposed. First, rather than repeating the acronyms in the target speech, interpreters are expected to deliver the Chinese names of the organisations, meaning that they would resort to their knowledge stored in LTM. If they are unfamiliar with the acronym, such as "*OECD*", the unfamiliarity would be reflected in retrieval time and effort (Cohen and Squire, 1980; Schacter, 1987; Baddeley, 2009). Also, some acronyms, such as "*UNICEF*", are perceived and pronounced as words. In fact, "*UNICEF*", in the SUBTLX-UK database (Hueven *et al*., 2014), has a Zipf value of 2.73, meaning that it is a rare word and that it would cost interpreters more time and WM resources to process than a word with higher frequency.

3.4 Conclusion

3.4.1 Summary of study

This corpus study is set out to investigate interpreters' GA during CI active listening and note again. Three RQs were asked, with the RQ1 concerning the overarching aspects of GA and the other two RQs tapping into the more specific aspects of the behaviour.

RQ1 asks the overarching question if interpreters would avert their gaze from note pads during notetaking. The results suggested that while interpreters did conduct GA, the behaviour took up a small proportion of total speech length. Also, GA is subject to considerable individual differences in terms of frequency and the time spent in GA.

RQ2 concerns whether interpreters would sustain or detach from taking notes during GA. Inferential analyses suggested that more GA was accompanied by the former, although the tests did not find significant differences between the two types of GA.

RQ3 examines textual features proceeding GA with the aim of narrowing down possible GA inducers. Altogether, seven textural features that could potentially increase interpreters' cognitive load were observed. The features either appeared alone or in combination. So far, the most commonly associated with GA were LTW and complex sentence structures. Quantitative data suggested that complex sentence structures (sentences with subordinate clauses) and LFWs seemed to be more powerful than the other features. A scoring system was developed where the sentence difficulty is defined by either complex sentence structure or LFW, or both. Inferential analyses suggested that GA sentences (sentences where GA occurred) were significantly more difficult than their counterparts, indicating that interpreters are significantly more likely to perform GA when they experience difficult sentences than easy sentences.

3.4.2 Strengths and Limitations

To the author's best knowledge, this study constitutes one of the earliest attempts to explore GA in interpreting studies via a mixed-methods design that combines the construction and coding of a corpus, a scoring system for sentence difficulty, and the observation of textual features as potential GA inducers. The study has shown strengths in two regards. First, the study made original contributions to knowledge by demonstrating that GA is associated with textual features that associated with high cognitive load, a finding that not only expands the range of application of GA from conversational settings to interpreting studies but also enriches the understanding of how interpreters' cognitive process can be embodied in their behaviours. Second, the methodology adopted in this study, especially the use of ELAN as a tool to code interpreters' behaviours, is likely to generate referential significance for studies with similar aims.

The study also entails shortcomings. As an effort to investigate eye movements, the study did not involve professional tools, such as an eye-tracker, to obtain relevant data, meaning that subtle GA, if any, was not coded. However, the absence of eye trackers was accepted as an inherited limitation not only because the videos were previously produced but also because the general aim of the study is to establish the fact that interpreters would look away from notepads so that the behaviour can be studied more rigorously in a follow-up study.

Chapter 4 The Experimental Study

This chapter represents the experimental study where findings from the previous chapters were verified in a controlled environment where all participants were exposed to the same speech. Also in this chapter, the intricacy of GA was explored further. In specific, the experimental study investigated whether exposing interpreters to LFWs contained in the upcoming speech would create a priming effect. Also, GA's potential performance-enhancing benefit was ~~also~~ explored. The methodological approach is presented in 4.1 where a research design informed by research aims and RQs is presented. Results are reported in 4.2. Discussion of results is located in 4.3. The chapter ends at 4.4, where a short summary of study is presented.

4.1 Methodology

4.1.1 Research Aim and Research Question

4.1.1.1 Research aim

The general aim of this experimental study encompasses two aspects: to verify the findings previously reported in Chapter 3 and to explore further the intricacy of GA in the setting of CI.

As previously mentioned, to the author's best knowledge, the corpus study constitutes one of the earliest effort in investigating interpreters' GA in the active listening and notetaking period of CI. Therefore, the first aim of the corpus was to establish that interpreters would resort to GA during the period so that an avenue could be paved for more in-depth exploration. The corpus study has shown three main findings: (1) Interpreters would avert their gaze from note pads at some point during active listening and notetaking; (2) During GA, interpreters would either fully detach from or remain engaged in the action of notetaking, but the difference between two types of GA did not reach statistical significance; (3) GA level, measured by GA frequency and time spent in GA, was positively associated with source speech containing complex sentence structure and/or LFWs. Albeit yielded ground-breaking findings, the corpus study was limited by some shortcomings: the source speech was not controlled in terms of topic, speech rate, the speaker, speech length, and difficulty; Also, the selection of source material (exam videos) for coding was limited by the video quality. Therefore, the findings from Chapter 3 would need to be verified in a controlled environment where participants are subject to consistent experimental conditions.

The second aim of the experimental study is primarily to explore the intricacy of GA further. Specifically, two aspects are encompassed in the second aim, and the first concerns whether manipulating interpreters' cognitive load by priming interpreters with LFWs in the upcoming speech would affect the level of GA. As previously discussed in 2.3.2.2, GA indicates cognitive load in a way that an increased level of cognitive load would result in a higher level of GA (Glenberg *et al.*, 1998; Doherty-Sneddon *et al.*, 2002; Doherty-Sneddon and Phelps, 2005). Based on this indicative role of GA, it seems justified to raise the question of whether a decrease in cognitive load would reduce ~~the~~ GA levels.

The second aspect pertaining to the second aim of the experimental study concerns whether GA would bear a performance-enhancing benefit in CI settings. As mentioned in 2.3.2.2, the Cognitive Load Hypothesis (Glenberg *et al.*, 1998) suggested that GA helps disengage from distractions and facilitates channelling more attentional resources to the task at hand. Disengaging from distractions would optimise the allocation of cognitive resources, contributing to better cognitive performances. Specifically, Glenberg *et al.* hypothesised that there is a positive correlation between the frequency of GA and cognitive performance. Phelps, Doherty-Sneddon, and Warnock (2006) taught young children to perform GA while thinking about answers to questions and found that training children to conduct GA during cognitively demanding tasks would tremendously improve concentration and task performance. However, in a study where participants with cognitive disorders went through face-to-face questioning, the authors reported that although participants' GA level was significantly influenced by task difficulty, the increased amount of GA did not show any effect on accuracy (Doherty-Sneddon, Riby, and Whittle, 2011). Despite the inconsistent findings from these studies, it has been clearly shown that people would resort to GA when cognitive load is high. As a ground-breaking study, to thoroughly investigate GA in interpreting studies, it would be crucial to test whether GA would bear any performance-enhancing benefit in CI settings.

4.1.1.2 Research Questions and Hypotheses

To adequately pursue the above research aims, the following RQs are developed:

> *RQ 1: Would the findings yielded from Chapter 3 sustain in the experimental study?*
> *RQ2: Would priming interpreters with LFWs contained in the upcoming speech result in a reduced level of GA?*
> *RQ3: Would GA enhance interpreters' performance?*

For RQ1, it is hypothesised that the findings from the corpus study will be verified in the experimental study. Specifically, interpreters will perform GA, and the GA levels would be significantly more prominent when the speech difficulty is higher. Also, regarding if GA would be accompanied by the action of notetaking, it is hypothesised that, like the corresponding finding reported in the previous chapter, interpreters would perform both GA+NT and GA-NT, with the former occurring significantly more frequently than the latter when the speech difficulty level is low. However, when the level of speech difficulty becomes high, GA-NT would be significantly more prevalent than GA+NT.

For RQ2, it is hypothesised that exposing interpreters to the upcoming rare words in the source speech would reduce processing load, leading to a reduced level of GA as WFE for those words would have been weakened by the exposure.

For RQ3, it is hypothesised that the performance-enhancing benefit of GA would manifest in the experimental study. Specifically, interpreters with higher GA levels would outperform those with lower GA levels.

4.1.2 Research Design

4.1.2.1 Piloting Stage

Addressing the above RQs posed several methodological challenges, and the primary one concerns the source speech, which should expose participants to an environment where desired variables are manipulated while interfering variables are controlled. The source speech should ideally have an equal proportion of independent variables so that the results and analysis would be credible. The balancing act also applied to speech length as it should be long enough to generate data without wasting valuable time. Also, considering the small pool of candidates, the suitability of tools and procedures would need testing, adjusting, and rehearsing to avoid data nullification. Additionally, the design of the experiment coincided with the Covid-19 pandemic. In the United Kingdom, where the experiment was conducted, the lockdown and social-distancing rules made it difficult to collect data. Therefore, the piloting stage allowed all factors to be considered before the experiment was initiated.

4.1.2.1.1 An Eye-Tracking Study

As mentioned in 2.3.1.2, although the eye-tracking technique began to emerge in interpreting studies in recent years (e.g., O'Brien, 2007; Seeber, 2012, 2013; Seeber and Kerzel, 2012;

Stachowiak-Szymczak, 2019; Tiselius and Sneed, 2020; Lin, 2022), the approach is well-grounded in the gaze-comprehension model which hypothesised that the mind comprehends visual inputs (Just and Carpenter, 1980). For studies that focus on GA (e.g., Doherty-Sneddon *et al*., 2012), eye-tracking technology is commonly used so that the qualitative features of GA can be studied in detail. Eye tracking data is usually obtained via an eye tracker that collects participants' gaze data at a specific frequency. For this study, a head-mounted eye tracker, such as SMI Glasses 2.0, which was used in Tiselius and Sneed's (2020) study, would be ideal as desktop eye trackers (e.g., Tobii X120) would require the participants to look at a monitor, which is not a standard notetaking set-up.

However, regrettably, the opportunity to use an eye tracker was eliminated by the social distancing rules forbidding people from different households from engaging in face-to-face interactions. Therefore, creative alternatives must be found to allow the collection of gaze data without involving social interaction with the participants.

4.1.2.1.2 Experiment Tools

The experiment involved two types of tools, and the first one concerns the collection of gaze data. Ideally, the tool for collecting gaze data should meet a few criteria. First, it should be easily accessible within the participants' living environment so that the social distancing rules would not be breached. Second, the tool must record participants' eye movements clearly and continuously. Third, the tool should be easy to use or involve minimal training so that the cost of time imposed on the experiment would be economical for all parties. Finally, the tool must allow straightforward data transmission from the participants to the author. It will be warranted to argue that smartphones with reasonably advanced front cameras would meet the criteria. Therefore, the tools used to collect case data were participants' smartphones. To avoid breaching social distancing rules, one of the criteria for recruiting participants was that they must have access to a smartphone that supports filming from the front camera.

Another tool involved in this experiment was Livescribe Echo Smartpens and Livescribe notebooks. The Livescribe Echo Smartpen is a sound and note recorder. Together with the notebook, the Smartpen can record the source speech and the process of notes production in a time-locked manner. Once plugged into a computer, recorded audio and notes can be accessed using the Echo Desktop application that can simultaneously play the process of notetaking and source speech. Given the inclusive recording function, Livescribe Echo Smartpens were used in some studies on CI notetaking (see Orlando, 2010; Chen, 2018). The author used the

Smartpen and the notebook during his MA project (Guo, 2016) and has gained adequate knowledge about generating and collecting data with the tools.

To avoid direct contact with the participants and to abide by social distancing rules during the COVID-19 pandemic, the Smartpens and notebooks were sent to participants by post. Before posting, pens and notebooks were sanitised. A pre-paid Royal Mail First Class postage label was included in each parcel, and the participants were asked to pack the Smartpen and notebook before dropping off the parcel at the local post office. Three Smartpens were rotated experiment. To avoid the overlap of notes produced by different participants, each participant's data was deleted from the Smartpens before the pen was posted to the next participant. None of the participants had experience with the notetaking duo. Therefore, an online training session was arranged after the Smartpens were delivered.

4.1.2.1.3 Variables

4.1.2.1.3.1 Independent Variables

Three independent variables were involved in this experiment: complex sentence structure, LFWs, and repetition priming.

Complex sentence structures were used to increase speech difficulty on the syntax level. The cognitive cost of processing complex syntax is well documented in the literature. As previously discussed in 2.6.2.4, active listening requires a listener to mentally compute the syntax of the sentence (Miyake *et al*., 1994; Chomsky, 1996; Caplan *et al*., 1998; Gibson, 2000; Wingfield and Tun, 2007; Frazier, 2013). The cost of such mental computing is found to positively correlate with the level of syntactic complexity. As reported in Chapter 3, interpreters' GA levels soared when complex sentence structure was experienced. To fulfil the first aim of this experimental study, complex sentence structures were used as independent variables.

The second independent variable was LFWs. As reported in Chapter 3, low word frequency was another powerful element associated with ~~an~~ increased ~~in~~ GA level. According to scholars, processing words triggers a lexical retrieval process where one must search the mental lexicon where low-frequency words are processed slower (Forster, 1976, 1981; Scarborough *et al*., 1977; Mutter and Hashtroudi, 1987; Keuleers *et al*., 2010; Brysbaert *et al*., 2018; Vitevitch *et al*., 2018; Neville *et al*.,2019; Popov and Reder, 2020). To test the finding from the corpus study, it would be necessary to include LFWs as another independent variable. However, the selection of LFW should be careful as some words, despite low frequency, are easily recognised (Brysbaert *et al*., 2018). For instance, the word "supersensitive" has a Zipf value of 1.47, making it a word that is rarely seen, and yet it can be easily recognised and understood by decomposing it into "super" and "sensitive". Also, despite low frequency, some words tend to be remembered upon the first encounter. For example, "Voldemort" (Zipf = 2.44) and "Dumbledore" (Zipf = 2.5) are rarely seen, but they would not pose a challenge for most *Harry Potter* fans. Therefore, the selection of LFW should avoid disguised rare words.

To investigate RQ2, repetition priming was introduced as the third independent variable. According to scholars, when a stimulus is experienced repetitively, the processing of relevant stimuli is more efficient (Schacter and Buckner, 1998; Rugg, Mark, Walla, Schloerscheidt, Birch, and Allan, 1998; Baddeley, 2009). Priming effects for LFWs are found to be stronger than common words (Forster and Chris, 1984). As mentioned in 2.6.2.4, handling LFWs is susceptible to WFE, where rare words consume more cognitive resources and time to process (e.g., Forster, 1976, 1981; Keuleers *et al*., 2010; Brysbaert *et al*., 2018; Vitevitch *et al*., 2018; Popov and Reder, 2020). Exposing interpreters to LFWs, therefore, would arguably clear a pathway for more efficient lexical retrieval, contributing to the possibility of reducing the consumption of cognitive resources during speech comprehension.

Given the significance between cognitive load and interpreting quality (Gile, 2009), investigating if a priming effect would occur would bear significance for interpreter training and practice.

4.1.2.1.3.2 Dependent Variables

As indicated by the aims of this experimental study, the dependent variables were primarily inherited from the previous chapter for the first two RQs. Specifically, total GA counts, GA length, GA counts per 100 source speech words, GA counts per 100 source speech seconds, GA proportion (time spent in GA to the total speech time), and GA type were measured and compared. For RQ3, which concerned whether GA would boost interpreters' performance, interpreters' performance score was also added as a dependent variable.

4.1.2.1.3.3 Controlled Variables

As previously discussed, one of the shortcomings of the corpus study was that a range of speech-related variables was not controlled, which is one of the main reasons why the findings of the previous chapter would need to be verified in this experimental study. To create an environment where the independent variables would be effective, the controlled variables included speech rate, speech length, the topic of the speech, and the speaker.

Studies have reported that listening comprehension would become less successful when the speech rate climbs (Conrad, 1989; Griffiths, 1992; Goh, 2000) and that a speech rate between 120 and 190 words per minute (wpm) has been considered a reasonable average delivery speed for speech used for comprehension (Marslen-Wilson, 1973; Riding and Vincent, 1980; Rivers 1981). Conrad (1989) controlled the normal rate at 180 w.p.m. when exploring the influence of speech rate on listening comprehension. Given the similarity between active listening in CI and listening comprehension, the speech rate was set at 180 wpm.

The length of the source speech would need to be controlled for two reasons. First, the length of the speech should be long enough for participants' cognitive load to be kept at a certain level to allow GA to emerge. Second, the length should allow all variables to be adequately experienced so that a sufficient amount of data would be generated. However, if the speech is lengthy, it could either cause fatigue, which would be an unwanted variable that affects the quality of interpreting (Moser-Mercer, 2000), or frustrate participants, pushing them not to take the task seriously. As mentioned in 2.6.1, CI is done in segments, and the length of segments is not unified (Setton and Dawrant, 2016) and can even last up to 20 minutes (Diriker, 2015). In this experimental study, a chunk of source speech lasted 210 seconds (3.5 minutes).

A general topic was selected for composing this speech. The primary reason for selecting a general topic is that background knowledge plays an important role in interpreting competence (Gafiyatova and Pomortseva, 2016; Al-Jarf, 2018). Having a specific topic would mean that interpreters who have acquired relevant knowledge would have an advantage over other participants, which would damage the credibility of the data.

In the corpus study, participants from 2015 rendered a speech uttered by a speaker who spoke with an accent. According to McAllister (2000), an unfamiliar accent can influence the quality of interpreting. Therefore, to ensure that accent would not pose as a variable, the speaker would ideally be a native English speaker who articulates well and uses Received Pronunciation.

Finally, the material used in the study, i.e. the source speech, was also controlled in terms of difficulty. The following section is dedicated to elaborating on the composition and the controlled difficulty.

4.1.2.1.4 The Source Speech

As previously mentioned, given the requirements for the source speech to be suitable for this empirical study, a source material that meets the criteria would need to be composed. To adequately explore the RQs raised in this empirical study, the source material included one easy speech and one difficult speech. Hypothetically, the discrepancy regarding speech difficulty would result in lower and higher levels of cognitive load, which would manifest in GA levels.

The easy speech, uttered in English, contained two chunks, with each chunk lasting 3.5 minutes and covering 630 words. The easy speech is composed of syntactically simple sentences and contained no LFWs. The difficult speech, also uttered in English, had the same number of chunks, length, and word counts. Compared with the easy speech, the difficult speech contained three types of sentences: complex sentences without LFWs, complex sentences with LFWs, and simple sentences with LFWs. Each chunk is composed with a mixture of three types of sentences, with each type accounting for a third of the word counts (210 words for each sentence type in each chunk). Hypothetically, participants would experience varying levels of cognitive load as a result of experiencing different speech difficulty, and GA measured under each difficulty level would be compared and analysed.

The easy and difficult speech was themed on “Hunger” and “Trust”, respectively. Both topics are general and can be easily understood. Under each speech, the two segments were not linked by a particular logic so that the order of speech and the order of chunks could be counterbalanced.

4.1.2.1.5 Priming

As mentioned before, priming was introduced as an independent variable. The rationale is that the participants primed with the LFWs would show a priming effect manifested in reduced GA levels and higher hit rates. SUBTLEX-UK (Hueven *et al*., 2014) was the database where

LFWs were selected. In total, 52 LFWs were chosen from the database and spread evenly into two segments, meaning that each chunk of the difficult speech contained 26 LFWs. The LFWs of each chunk were extracted and organised in a glossary list (see appendix 3) containing the following information:

- The LFW
- Phonetic symbols (pronunciation)
- The contextual definition in Chinese
- An example sentence offering a similar context to the source speech
- Chinese translation of the example sentence.

To ensure homogeneity, 52 words were chosen from the easy speech. Given the high word frequency of the easy words, it is likely that the participants would have already stored them in their mental lexicon. Therefore, the easy words were listed without Chinese definitions or context.

Table 9 below shows an example of the word list from easy and difficult speech.

Table 9 Example of glossary list from easy and difficult speech

Easy Source speech	***Difficult Source Speech***
Poorest	Hanker [ˈhæŋkə] 渴望 Example Sentence: We all hanker for love. Translation: 我們都渴望愛。
Opposite	ruminating [ˈruːmɪneɪtɪŋ] 思考 Example Sentence: I have been ruminating the issue of poverty. Translation: 我一直在思考貧困這個話題

4.1.2.1.5 Pilot Study I

The first pilot experiment was conducted to fulfil two aims: to test if the experiment design fits the purpose of the experimental study; to rehearse the procedure so that issues can be flagged and solved for the experiment.

Participant

One participant (female, 26 years old) was recruited for this pilot study after consent was obtained. She is a native Chinese speaker with advanced English proficiency as she works as a professional interpreter between the two languages. The participant was rewarded a MA's degree in Interpreting from Newcastle University in 2016.

Procedure

The pilot experiment involved several tasks and was conducted online. Since the author could not be there to set up the experiment, the participant had to follow manual instructions. It is necessary to report that the source speech was played in the form of audio by the author. The sound was transmitted to the participant via the built-in microphone of a computer running a MacOS. To give an exhaustive description, the procedure is described in a step-by-step manner below:

- Step 1: Briefing. The participant was briefed about the overall aim of the study. Queries from the participant were clarified, but details that could potentially influence results were not given.
- Step 2: Camera preparation and testing. The participant was told to set the smartphone to Flight Mode but keep a Wifi connection so that a Microsoft Teams call could be made from the gadget. The participant was then instructed to turn the phone upside down so the camera could record the interpreter's face from a bottom-up angle. The angle of the phone was adjusted until the author was confident that eye movements could be clearly captured.
- Step 3: Testing Smartpen and Livescribe Notebook. A demonstration left using the notetaking duo was shown to the participant. After the demonstration, the participant was given 2 minutes to practise using the duo.
- Step 4. Warm-up. A 2-minutes pre-recorded warm-up speech by the same speaker was played. The main purpose of the warm-up session was to give the participant a chance to get used to the speaker's voice and speech rate. It was also an opportunity for the participant to practise taking notes with the tools.
- Step 5. A word list containing 26 words from the easy speech was shown on the screen via screen-sharing function. The participant was instructed to go through the list once.

- Step 6. Play the audio of the first chunk of the easy speech.
- Step 7. Interpret the first chunk.
- Step 8. Play the audio of the second chunk of the easy speech.
- Step 9. Interpret the second chunk of the easy speech.
- Step 10. Prime the participant with 26 LFWs from the first chunk of the difficult speech via the sharing screen function of Teams. All 26 words were listed in one screen with easy-to-read font size. She was instructed to familiarise herself with the words in her own way, but inventing symbols was not allowed. At the end of the priming session, the participant was instructed to close the file.
- Step 11. Play the audio of the first chunk of the difficult speech.
- Step 12. Interpret the first chunk of the difficult speech.
- Step 13. Play the audio of the second chunk of the difficult speech.
- Step 14. Interpret the second chunk of the difficult speech.
- Step 15. Remind the participant to turn off the Smartpen and announce the completion of the experiment.

Issues and Adjustments

A few issues surfaced from the first pilot study. First, during the experiment, the participant wore ear pods, which were overlooked. As a result, the audio of the source speech was not picked up by the Smartpen. For the second pilot experiment, the participant would need to be reminded not to use headphones.

Second, only priming interpreters with one list of LFWs seemed to be inadequate. According to the feedback from the participant, she barely had any impression of the words that she was primed with. For the other 26 LFWs she was not primed with, she did not know them at all. As a result, although the participant did make an effort to recognise the words, as soon as she realised that they were not included in her vocabulary, she did not attempt to conduct lexical retrieval. To respond to the second issue, The author decided to alter the priming/no-priming approach to longer priming (8 minutes)/shorter priming (4 minutes) approach.

Third, the participant reported that the 180 wpm speech rate was excessively fast for CI. The speech rate, therefore, was dimmed down to 140 wpm, meaning the word count for each chunk would be 490. To maintain an equal proportion of each type of sentence in the difficult speech, the word count for each chunk was set at 489.

Table 10. below shows an example of the adjusted source speech.

Table 10 Example of Sentences from Easy and Difficult Source Speech

Easy Source speech	***Difficult Source Speech***
A few years ago, developed countries made a donation. They gave money to the poorest countries. The goal was to solve one of the world's biggest problems.	However, my mother <u>whom</u> the harshness of life has **inured** watched the video with me and **eulogised** the show <u>as what</u> she believes is <u>that</u> <u>it is</u> paramount to remind people <u>that</u> our time is defined by **treachery which** is unseen by any other times.

Readability Check

The readability of each source speech chunk under both difficulty level was checked using Microsoft Word's Readability checker. The function adopts Felsch Reading Ease scale. According to Stockmeyer (2009), a minimum Felsch readability score of 60 would indicate the document is composed with plain English, and authors should aim for a score between 60 and 70 to ensure good readability. The results of the readability suggested that all speech chunks should be understood with ease because the readability scores for all four chunks were within the range between 60 and 70. In specific, the scores for the two chunks under low speech difficulty were 68.4 and 70.1, respectively, whereas the scores for the two difficult speech chunks were 63.4 and 66.4, respectively.

4.2.1.2.6 Pilot Study II

The second pilot experiment was conducted to fulfil two aims: to test if a counterbalanced speech order will result in contextual confusion; to test if the updated speech rate, length, and priming protocol would increase the suitability of the experiment design.

Participant

Another participant (Female, 28) gave consent to participate in the study. The participant is a native Chinese speaker with advanced English proficiency. She holds an master's degree in Translation and Interpreting, and use both Chinese and English as working language. The second pilot study used the updated source speech and adjusted procedures regarding GA.

Procedure

The procedure was almost identical to the first one with two main differences:

- The speech order was reversed to test if there was an order effect. Specifically, instead of starting with the first chunk of the easy speech, the participant in the second pilot study began her rendition from the second chunk of the difficult speech.
- Priming procedure – a list of the upcoming words was provided before each chunk was uttered. For the easy speech, the participant was told to go through the list once. For the difficult speech, the participant was given 4 minutes for one list before the chunk was uttered and 8 minutes for another list before the recording of the corresponding chunk was played.

Issues and adjustment

As previously discussed, the source speech was composed in a way that the two chunks did not follow a specific logic. According to the participant, the two chunks were coherent with a natural flow of information. Also, the longer and shorter priming sessions seem to have given a chance for the participant to gain some familiarity with the words, as some of the words were faithfully rendered in the target speech. The second pilot study showed that the previous adjustments were effective and did not indicate any new major issues.

4.2.1.2.7 The Experiment

Participants

The institution where the research is conducted had previously granted Ethical Approval to this study to recruit human participants in face-to-face experiments. However, due to the adverse circumstances imposed by COVID-19, conducting face-to-face activities would break social distancing rules from both the government and the university. A modification of the methodology was carried out, and as a result, the experiment was conducted remotely on Microsoft Teams. The change of methodology required another ethical approval, which was obtained prior to the experiment.

The participants, all female, aged between 22 and 25, were 16 continuing Class 2021 students from the MA Translating and Interpreting Programme of Newcastle University. All participants are native Chinese speakers. None of the participants had previously acquired work experience as an interpreter. However, by the time of the experiment, the participants had studied CI as a core module for three semesters and had adequate practice with CI notetaking both in and out of classes. Also, they all passed Final Exam for CI for Stage 1 students. To ensure that all participants are equipped with a similar level of expertise in CI and to limit individual differences regarding capability, only those who had scored 60 and above in the final exam of CI were contacted. This score is the threshold for selecting students to continue their studying of the CI module in the second year of the programme. Compared with their counterparts who scored under 60, the participants demonstrated good skills and capabilities for satisfying CI performances.

An information sheet was sent to all potential participants to briefly introduce the purpose and procedure of the experiment. However, instead of revealing that their eye movements would be studied, they were told that the purpose of the study was to explore cognitive load during CI. Participants were encouraged to email the author their queries regarding the experiment. All issues were clarified before a consent form was sent to the participants.

Procedure

The procedure of the study copied the adjusted protocol used in the pilot study (see 4.1.2.1.5 and 4.1.2.1.6). To avoid repetition, please refer to sections 4.1.2.1.5 and 4.1.2.1.6 for a detailed description. However, it is important to note that the 16 participants were divided into four groups, with each group containing 4 participants. To avoid order effects, the order of chunks of each speech and the order of two speech was counterbalanced. In terms of priming, interpreters from Group 1 and Group 2 were given 8 minutes to gain familiarity with the first list of LFWs and 4 minutes for the second list of LFWs. By contrast, interpreters of Group 3 and Group 4 were given 4 minutes for the first LFW list 8 minutes for the second LFW list.

4.1.3 Data Collection, Coding, and Marking

4.1.3.1 Data Collection and Processing

In this experimental study, two types of data were collected from two tasks: eye-tracking and CI. Collecting gaze data was achieved by following the protocol laid out in Chapter 3, where an exhaustive step-by-step procedure was provided (see 3.1.2.3). Data from the CI task involved interpreters' rendition of the source speech and CI notes. The recording of the rendition was readily available as the whole experiment process, including interpreters' voices, was recorded. Obtaining CI notes, owing to the user-friendly Livescribe Desktop application, was straightforward.

4.1.3.2 Coding GA

Interpreters' GA was coded in ELAN. As previously mentioned, the source speech used in the experimental study contained independent variables. To accurately reflect the variables in mark-ups, a new set of codes were used. See Table 11 below.

Table 11 Mark-ups and definitions

Code	Description
GA+NT(SS)	GA accompanied by the action of notetaking during processing simple sentences from the easy speech.
GA-NT(SS)	GA not accompanied by the action of notetaking during processing simple sentences from the easy speech.
GA+NT(CS)	GA accompanied by the action of notetaking during processing syntactically complex and LFW-free sentences from the difficult speech.
GA-NT(CS)	GA not accompanied by the action of notetaking during processing syntactically complex and LFW-free sentences from the difficult speech.
GA+NT(LFW)	GA accompanied by the action of notetaking during processing syntactically simple sentences that contain LFWs from the difficult speech.
GA-NT(LFW)	GA not accompanied by the action of notetaking during processing syntactically simple sentences that contain LFWs from the difficult speech.
GA+NT(Both)	GA accompanied by the action of notetaking during processing syntactically complex sentences that contain LFWs from the difficult speech.
GA-NT(Both)	GA not accompanied by the action of notetaking during processing syntactically complex sentences that contain LFWs from the difficult speech.

The coding procedure rigorously mirrored the protocol laid out in the previous corpus study. To avoid repetition, please refer to 3.1.4.2.4.

Scholars suggested that inter-rater reliability checks are necessary for corpus-based studies to ensure the accuracy of coding (Biber *et al*., 2007; Connor and Upton, 2007; Upton and Cohen, 2009). To abide by the standard, the coder who performed the inter-rater reliability check for the corpus study gave her consent to conduct the task in the experimental study. The inter-rater (female, aged 27), is a native Chinese speaker studying in Newcastle University as a PhD candidate. She as a master's degree in Linguistics from a university in the UK and has been using English and Chinese as working languages for years. Given the time elapsed between the two tasks, the coder went through a training session during which the definitions and protocol of annotations were provided in written forms, which were followed b explanations and clarifications. During the session, the coder practised adding and adjusting annotations in ELAN. There were some discrepancies regarding the length of some GA, but the differences are minimal, and most of the differences were reconciled during the post-coding discussion. The inter-rater reliability check rendered a 98% agreement, with the disagreement reconciled through discussion.

4.1.3.3 Marking

To test whether GA enhanced interpreters' performance (RQ3), the target speech from all participants were marked. The performances of interpreters in this experimental study were measured by completeness and accuracy. Completeness would be damaged by omissions of information, whereas accuracy would be jeopardised when a piece of information was not accurately rendered. Considering that sentences are composed of smaller information units (Cutler and Clifton, 1999), examining completeness and accuracy with rigour would require source sentences to be segmented into smaller units. Bovair and Kieras (1981) proposed a set of meticulous principles for segmenting sentences for propositional analysis. For instance, following the principle, a source sentence "*I have to adapt to their working hours*" can be segmented into 6 information units "*I/have to/adapt to/their/ working/ hours*". The performance for this particular sentence would be based on the equation:

$$Performance\ Score\ (\%) = \frac{Number\ of\ faithfully\ rendered\ units}{Total\ number\ of\ information\ units}$$

Added information in the target speech, unless it distorted the meaning of the information units of a particular sentence, was not penalised. For example, if the above sentence was interpreted as "*She tried to learn from their working habit*", the rendition would have obtained a score of 33.33%, as only 2 units of information were faithfully rendered. However, if the rendition is "*I have to adapt to their working hours every day*", then the rendition would have achieved a full mark (100%) as all units were completely and accurately rendered. The total performance for a speech or a chunk is calculated via the following equation:

$$Total\ Score\ (\%) = \frac{Sum\ of\ scores\ for\ each\ sentence}{total\ number\ of\ sentence}$$

Inter-rater reliability check was also conducted for the scoring of performance. An inter-rater (Female, 31) gave her consent to check the reliability of the marking. The inter-rater is a native Chinese speaker studying in Newcastle University as a PhD candidate. She obtained a master's degree in Translation and Interpreting from Newcastle University in 2017 and has been working as a freelance interpreter while studying in the same institute as a PhD candidate. Before the task, the inter-rater went through a briefing where the task was explained. The performances of 4 interpreters were checked. An initial 96% of agreement was reached, and all discrepancies were reconciled.

4.1.4 Data Analysis

4.1.4.1 Descriptive Analysis

Similar to the corpus study, the collected data was initially organised in spreadsheets so that the patterns of data could be quantitively described or summarised. As previously shown in the corpus study, coding a corpus usually results in abundant and messy data (Stuart *et al*., 2016). Given the depth and scope of this experimental study, the abundance of data regarding GA would arguably be enriched even further. To organise the data in a way that facilitates straightforward reviews and descriptive analysis, a master spreadsheet listing all GA measures for each interpreter was created. Data regarding priming and interpreters' performance was also initially organised in spreadsheets so that the data pattern could be summarised and described in a qualitative manner. To show an example, Table 12 below is extracted from the master spreadsheet based on which data were manipulated and exploited to produce tables and diagrams for reporting data patterns in 4.2.

Table 12 Example of Master Sheet for GA under each speech

Interpreter	***Total GA Counts***	***Total GA Length (Seconds)***	***Average GA Length for***	***GA/100 Source***	***GA/100 Speech (Seconds)***	***GA proportion (%)***

			Each Interpreter	***Speech Words***		
1	0	0.00	0.00	0	0.00	0.00
2	1	1.44	1.44	0.10	0.24	0.34
…	…	…	…	…	…	…
16	18	10.44	0.58	1.84	4.29	2.49
Total	33	28.89	NA	3.37	7.86	NA
Mean	6.6	0.88	0.65	0.67	1.57	1.38
SD	8.71	7.64	0.67	0.89	2.07	1.82

4.1.4.1.1 Inferential Analysis

Scholars in quantitative studies (e.g., Stockemer *et al*., 2019; Bloomfield and Fisher, 2019) highlighted the significance of inferential analysis as it helps researchers to test if the observed differences between data groups reached statistical significance so that a more reliable conclusion can be drawn. In this experimental study, inferential analysis was performed after the data had been quantitatively observed and described. The assumption of normal distribution was tested for all data groups to help the author to make informed decisions when selecting an inferential test so that the differences between data groups can be robustly and accurately compared (Field, 2013; Gerald, 2018).

4.1.5 Summary

In summary, this experimental study was conducted to verify previous findings from the corpus study and to further investigate the intricacy of GA in CI settings. Tasks involved in this experimental study include an experiment, coding of GA, and marking interpreters' performance. The material used in the study consisted of an easy speech and a difficult speech. The easy speech did not contain rare words and adopted syntactically simple structures. By comparison, difficult speech contained three types of sentences: syntactically simple sentences containing LFWs, syntactically complex sentences without LFWs, and syntactically complex sentences containing LFWs. Priming was also involved in this study to see whether exposing interpreters to the upcoming LFWs would result in a reduced GA level. Interpreters' gaze data was coded in ELAN before being exported to a spreadsheet for quantitative and inferential analysis. Interpreters' performances, which were measured by completeness and accuracy, were also quantitively and influentially compared. The results obtained are presented in the next chapter.

4.2 Results

4.2.1 GA under Easy and Difficult Conditions

To reiterate, the first RQ pursued in this experimental study asked if speech difficulty defined by complex syntax, LFWs, or both will correlate with GA during CI active listening and notetaking. This research question is underpinned by previous findings that, during conversations, people would look away from distractions (e.g., interlocutor's face) when answering difficult questions to circumvent cognitive overload (Glenburg *et al*., 1998; Doherty-Sneddon *et al*., 2002). In this study, it is hypothesised that compared with the easy speech that contains only simple sentences and common words, Speech B which contains complex sentences, complex sentences with LFWs, and simple sentences with LFWs, will result in a higher GA level measured by counts and lengths because processing the difficult speech would impose a higher level of cognitive load. Within the scope of RQ1, it is further hypothesised that, within the difficult speech, sentences that contain the two difficulty variables would be most powerful in inducing GA due to the additive nature of cognitive load (Galy *et al*., 2012).

4.2.1.1 Overall Description

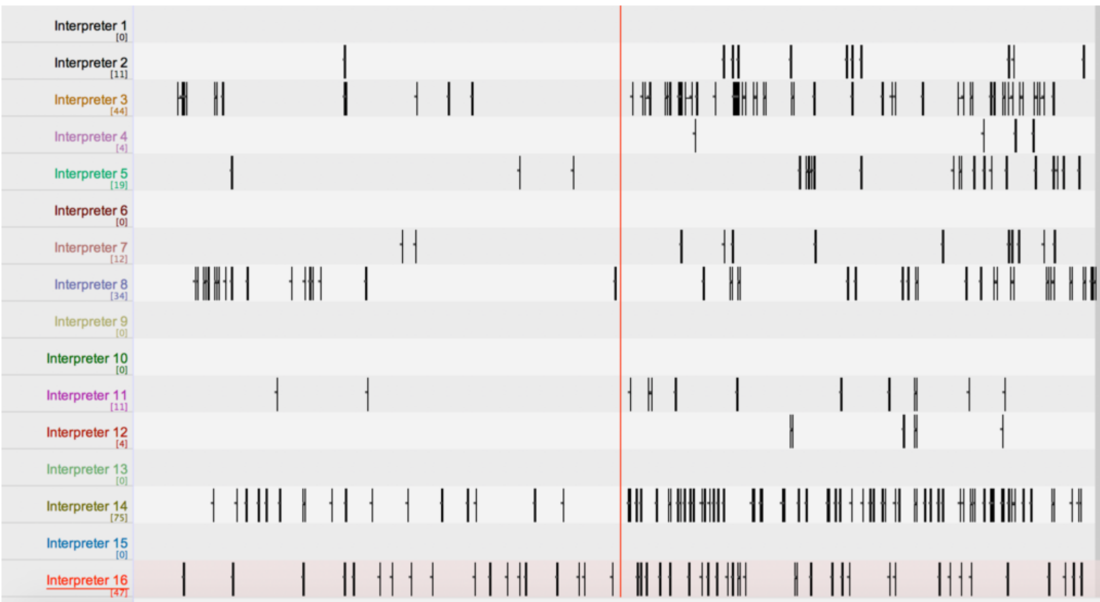


Figure 22 Mark-ups in ELAN in the experimental study

Figure 22 is the interface of ELAN after the coding of GA was completed. The red line across the middle indicates the boundary between the easy (left) and difficult (right) speech. An initial glimpse of the scatter of the mark-ups would indicate that the density of mark-ups is higher under the difficult speech condition than the easy speech condition, meaning that subjects averted their gaze more often in response to the increased speech difficulty. Another visible feature from the interface of ELAN would be some empty layers where GA was not coded. In specific, when the speech was easy, 8 participants performed GA. By contrast, when processing the difficult speech, 10 interpreters averted gaze from notepads during active listening.

As indicated in 4.1.3.2, mirroring the measures in the corpus study, GA in this experimental study was also measured in terms of frequency and dwelling time. In specific, the frequency measures include GA counts, GA counts per 100 source speech words, and GA counts per 100 source speech seconds, whereas the dwelling time measurements include GA length and GA proportion (total GA time to total speech length). A box plot (Figure 23) below is constructed to illustrate the distribution of data regarding GA measurements for all interpreters under the easy speech based on the master sheet (see Appendix H).

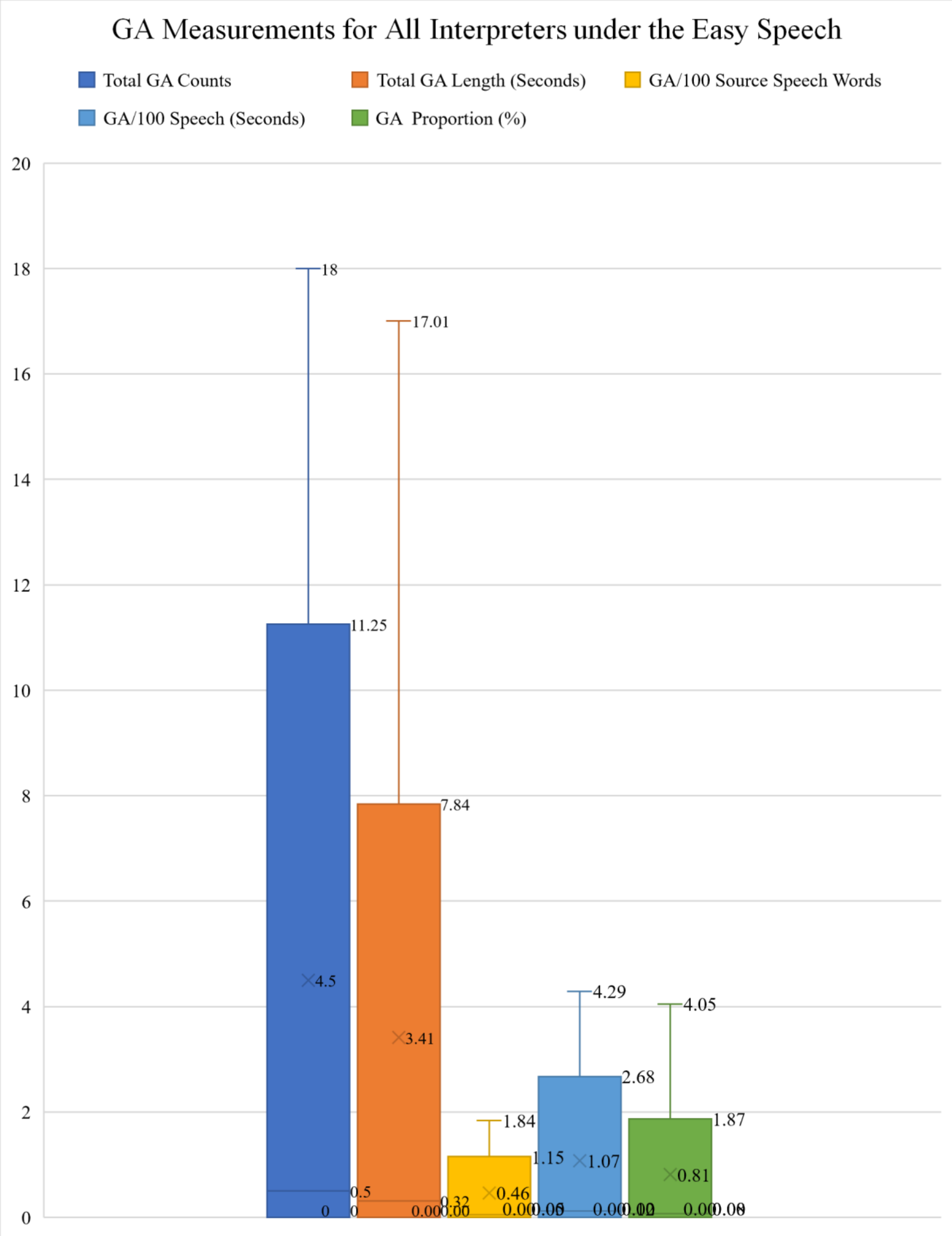


Figure 23 GA measurements for all interpreters under the easy speech

In terms of total GA counts, the median score is 0.5, with an interquartile range between 0 (Q1) and 11.25 (Q3), suggesting that half of the interpreters averted gaze from notepads under 11.25 times. On average, each interpreter performed 4.5 times of GA (*max* = 18, *min* = 0). Regarding total GA length, the median score is 0.32 seconds, with an interquartile range between 0 (Q1) and 7.84 seconds (Q3). On average, each interpreter invested 3.41 seconds in averting gaze from the notepad (*max* = 17.01, *min* = 0). The median for the score of GA counts per 100 words of the easy speech is 0.05. The interquartile range is between 0 (Q1) and 1.15 (Q3). On average, 0.46 times of GA were coded for every 100 words uttered by the speaker (*max* = 1.84, *min* = 0). For GA coded every 100 seconds of the easy speech, the median score is 0.12, with an interquartile range between 0 (Q1) and 2.68 (Q3). The average amount of GA coded per 100 seconds is 1.07 (*max* = 4.29, *min* = 0). Finally, pertaining to GA proportion, the median value is 0.08%, with an interquartile range between 0% and 1.87%, suggesting that half of the interpreters spent less than 7.85 seconds conducting GA while processing the source speech (*max* = 4.05%, *min* = 0%). Compared with the 2013 and 2014 cohorts, it is evident that the distribution of GA data obtained from the 2015 cohort is positively skewed with the maximum scores being outliers located outside the left whisker. It is evident that data distribution for all measurements is positively skewed.

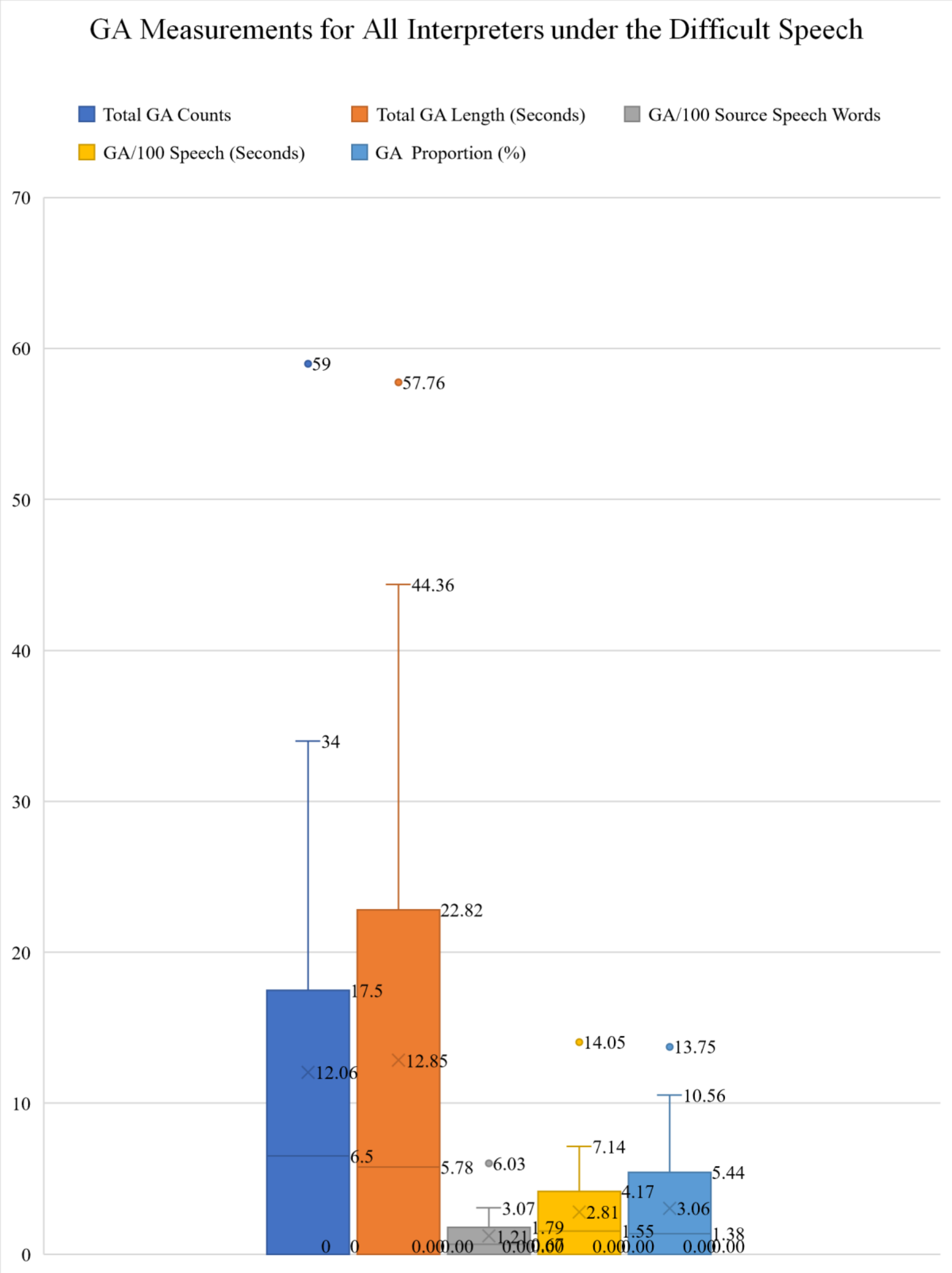


Figure 24 GA measurements for all interpreters under the difficult speech

Figure 24 above is plotted to display the distribution of GA measures pertaining to frequency and proportion per the master sheet (see Appendix H) for the difficult speech. In terms of total GA counts, the median score of total GA counts is 6.5, with an interquartile range between 0

(Q1) and 17.5 (Q3), indicating that half of the interpreters performed 17.5 times of GA or less. On average, each interpreter conducted 12.06 times of GA (*max* = 59, *min* = 0). Regarding total GA length, the median score is 5.78 seconds, with an interquartile range between 0 and 22.82 seconds. On average, each interpreter spent 12.85 seconds looking away from their notepads (*max* = 57.76, *min* = 0) while processing the difficult speech. For GA counts per every 100 words from the speech, the median score is 0.67, with a Q1 at 0 and a Q3 at 1.79. On average, 1.21 times of GA were coded for every 100 words uttered by the speaker (*max* = 6.03, *min* = 0). The median number of GA coded every 100 seconds of the source speech is 1.55, with an interquartile range between 0 (Q1) and 4.17 (Q3). On average, each interpreter performed 2.81 times of GA during each 100 seconds of the source speech (*max* = 14.05, *min* = 0). Finally, regarding the proportion of time spent in GA to speech length, the median score is 1.38%. The interquartile range is between 0% (Q1) and 5.44% (Q3), suggesting that half of the interpreters spent up to 22.85 seconds averting gaze from notepads. On average, 3.06% of the time (or 12.85 seconds) was invested in GA (*max* = 13.75%, *min* = 0%). Overall, data distribution for each data type seems to be positively skewed with an outlier located beyond the upper whisker.

Comparing GA measures between the easy and the difficult speech, based on the descriptive data, would indicate that the increased speech difficulty was associated with a higher GA level, paving the way for the difference to be inferentially tested.

4.2.1.2 Inferential Analysis

In line with the corpus study and based on the research design, inferential analysis was conducted to test whether GA under easy and difficult speech was significantly different. To ensure robust and adequate analysis, a Shapiro-Wilk test was performed to examine the assumption of normality. Table 13 below summarises the results of the normality checks. Based on the results and the 2-group feature, a non-parametric Wilcoxon sign ranked test was performed to test whether the difference regarding GA measures under easy and difficult speech conditions.

Table 13 Shapiro-Wilk Test Results for GA Measure under Easy and Difficult Speech

Data Group	***Shapiro-Wilk Test Result***	***Evidence for Departure from Normality***
Each interpreter's GA counts (easy)	$w(16) = .66$, p <.001	Yes
Each interpreter's GA counts (difficult)	$w(16) = .76$, p <.001	Yes
Each interpreter's Avg. GA length (easy)	$w(16) = .66$, p <.001	Yes
Each interpreter's Avg. GA length (difficult)	$w(16) = .76$, p <.001	Yes
GA/100 words (easy)	$w(16) = .78$, p <.001	Yes
GA/100 words (difficult)	$w(16) = .66$, p <.001	Yes
GA/100 seconds (easy)	$w(16) = .66$, p <.001	Yes
GA/100 seconds (difficult)	$w(16) = .77$, p <.001	Yes
GA proportion (easy)	$w(16) = .66$, p <.001	Yes
GA proportion (difficult)	$w(16) = .76$, p <.001	Yes

The Wilcoxon test revealed statistical significance across the groups. Regarding GA counts, the test revealed that interpreters performed significantly more GA under the difficult speech condition (Md = 6.50, n = 16) than easy speech condition (Md = .50, n = 16), z = -2.81, p = .005, r = .70. In terms of average GA length, the test found that GA last significantly longer when the speech was difficult (Md = 5.78, n = 16) than when it was easy (Md = .31, n = 16), p = .005, r = .70. For every hundred words from the source speech, the frequency of GA was significantly higher under the difficult speech (Md = .67, n = 16) condition than the easy speech condition (Md = .05, n = 16), p = .005, r = .70. Similarly, GA frequency measured by every 100 seconds of the speech length was also significantly higher when the interpreter processed difficult speech (Md = .1.55, n = 16) than easy speech (Md = .12, n = 16), p = .005, r = .70. Finally, compared with the easy speech (Md = .08., n = 16), GA proportion, or time spent in GA, was significantly higher in the difficult speech (Md = .1.38, n = 16), p = .005, r = .70. Figure 25 (below) is plotted to help visualise speech difficulty's significant effects over GA. In specific, for each measurement, the statistical significance between low and high speech difficulty is indicated by two asterisk marks (as p values are .005) and a significance line.

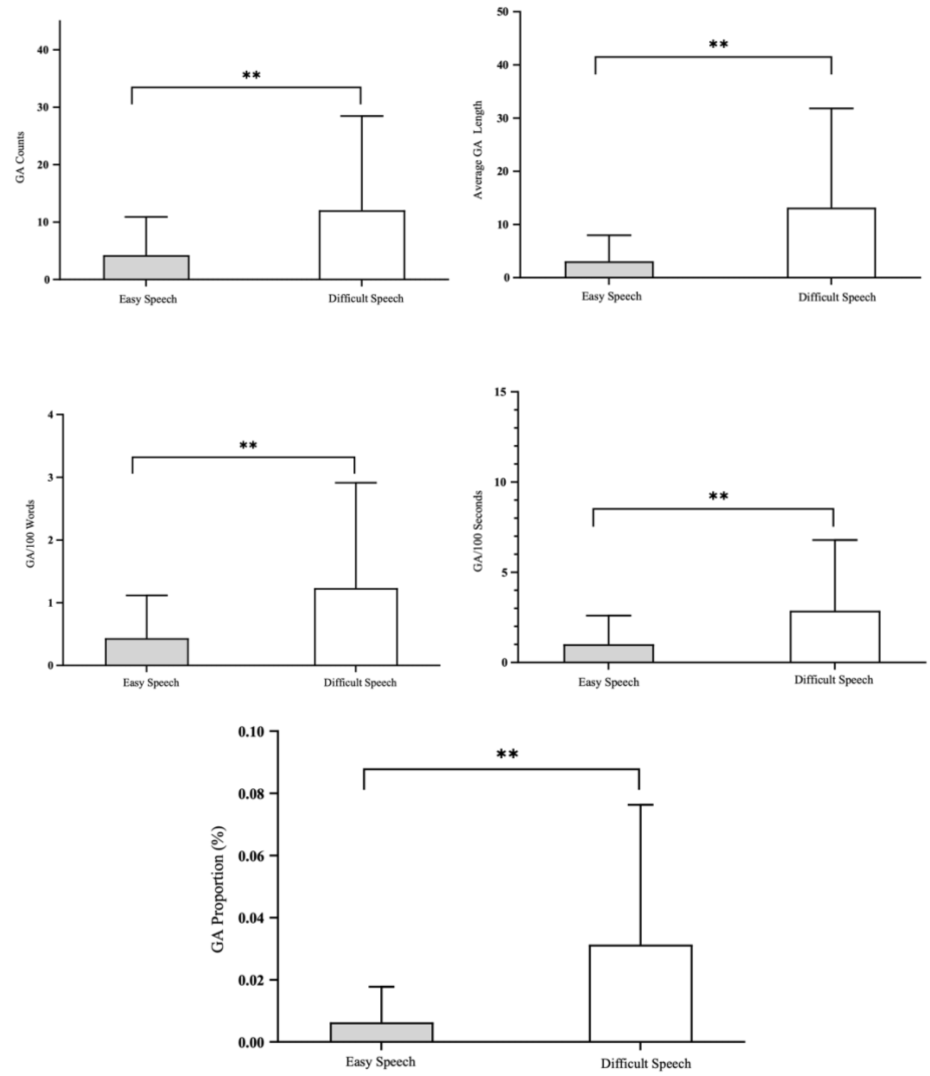


Figure 25 Significant effects across measurements

Based on the above results from descriptive and inferential analysis, it is arguably the case that despite drastic individual differences in GA measures, high speech difficulty imposed by complex sentence structure, or LFWs, or both, has a significant effect in affecting GA during CI active listening and notetaking. Specifically, during the task, an increase of speech difficulty would prompt interpreters to look away from notepads more often and spend more time averting gaze. Hypothesis 1 is, therefore, accepted. This finding is in line with previous studies where people were found to look away from distractions to cope with the climbing cognitive load (Glenberg *et al*., 1998; Doherty-Sneddon *et al*., 2002; Doherty-Sneddon *et al*., 2012). In other words, looking away from notepads during CI can be considered as an indicator of cognitive load, with counts and lengths rising as the result of increased cognitive load.

4.2.1.3 Summary

The overarching finding from the previous corpus study was interpreters would conduct GA during CI active listening and notetaking, and the level of GA would increase in response to a rise of speech difficulty. However, as reported in the previous chapter, GA levels are subject to considerable individual differences. In this experimental study, GA was also found among interpreters, which would verify the first aspect of the finding from the corpus study. However, unlike the previous finding where all interpreters performed GA in spite of dramatic individual differences, data showed that, in this experimental study, some interpreters did not resort to GA at all. For GA interpreters, the behaviour was found to be significantly more prominent across all measures when the difficult speech than the easy speech was processed, corroborating the other aspect of the overarching finding from the previous chapter.

4.2.2 GA in the Difficult Speech

4.2.2.1 Descriptive Data

In the context of the experimental study, it was also hypothesised that among the three levels of difficulty, sentences combined complex structure and LFWs would see the highest GA levels because the gross cognitive load from accommodating the two variables. To test the hypothesis, a zoomed-in view was casted at the GA obtained under the difficult speech condition by organising GA under each variable. The box plot (Figure 26) below is plotted to display the distribution of counts and average length of GA coded under each sentence type. A table that summarises the data above can be found in Appendix I.

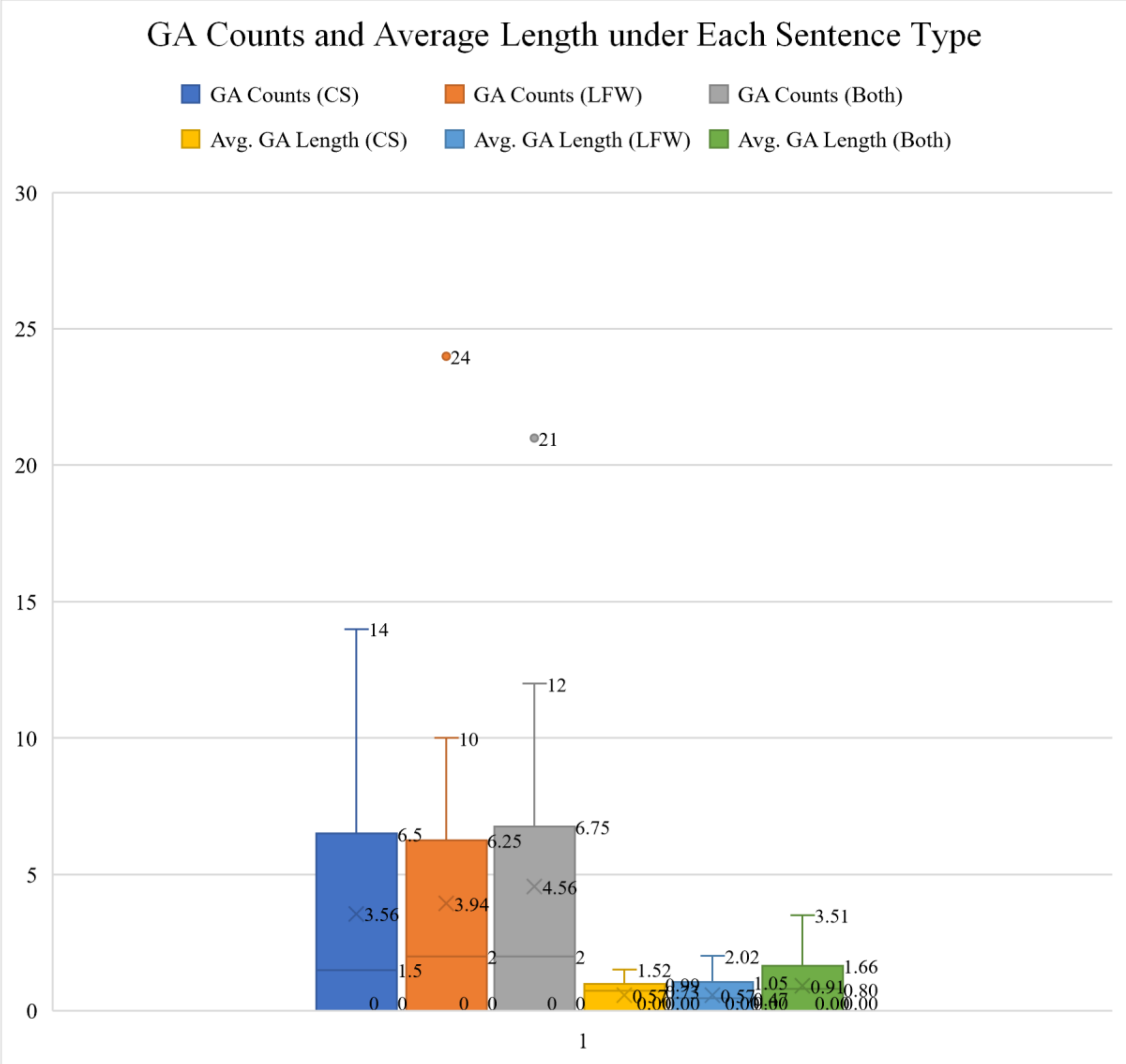


Figure 26 GA counts and average length under each sentence type

For sentences with complex structure, the median score for GA counts is 1.5, with an interquartile range is between 0 (Q1) and 6.5 (Q3), meaning that half of the interpreters performed less than 6.5 times of GA when processing the complex sentences. On average, 3.56 times of GA were coded (*max* = 14, *min* = 0). For simple sentences that containing rare words, the median score for the number of GA coded is 2. The interquartile range is between 0 (Q1) and 6.25 (Q3), with an average GA count of 3.94 (*max* = 24, *min* =0). Regarding GA coded under sentences with both variables, the median score is 4.56, covering an interquartile range between 0 (Q1) and 6.75 (Q3). On average, each interpreter performed 4.56 times of GA when processing complex sentences with rare words (*max* = 21, *min* =0). The median score for the average GA length under the complex sentence condition is 0.73 seconds, with an interquartile range between 0 (Q1) and 0.99 (Q3) seconds. On average, each interpreter spent 0.57 seconds averting gaze from their notepads (*max* = 1.52, *min* = 0). For simple sentences containing rare words, the median score for average GA length is 0.47 seconds. The interquartile range for this measurement is between 0 (Q1) and 1.05 (Q3) seconds. On average, each GA lasted 0.57 seconds (*max* = 2.02, *min* = 0). Finally, for sentences containing both variables, the median score for average GA length is 0.8 seconds, with an interquartile range between 0 (Q1) and 1.66 (Q3) seconds. On average, each GA coded under the condition of both variables lasted 0.91 seconds.

Based on the data from the table above, source sentences that combined two difficulty variables attracted more GA than the other two types of sentences as shown by both counts and length variables. To confirm whether the difference reached statistical significance, inferential analysis was conducted.

4.2.2.2 Inferential Analysis

Prior to selecting a robust test, the assumption of normality was examined via a Shapiro-Wilk test. Table 14 below is drafted to report the normality check results. The test revealed that the scores for all groups were not normally distributed, and therefore a non-parametric Friedman test was chosen to test the differences.

Table 14 Shapiro-Wilk test results for GA measures under three types of source sentences

GA Type	***Shapiro-Wilk Test Result***	***Evidence for Departure from Normality***
GA Counts (Complex Structure)	$w(16) = .83, p = .006$	Yes
GA Counts (LFW)	$w(16) = .68, p <.001$	Yes
GA Counts (Both)	$w(16) = .81, p = .004$	Yes
Avg. GA Length (Complex Structure)	$w(16) = .87, p = .028$	Yes
Avg. GA Length (LFW)	$w(16) = .83, p = .008$	Yes
Avg. GA Length (Both)	$w(16) = .88, p <.043$	Yes

The Friedman test did not render any significant effect among GA counts measures, $\chi2(2) = .1.8$, p =.41, or average GA length, $\chi2(2) = 5.6$, $p = .06$. The results suggested that there is no statistical difference regarding GA levels among three types of sentences in the difficult speech.

4.2.2.3 Summary

In a nutshell, under the difficult speech condition, there was no statistically significant difference between GA count measures and GA length measures among the three types of sentences.

4.2.3 GA+NT vs GA-NT

The second finding that this experimental study was set out to verify concerns the fact that interpreters from the corpus steady tend to perform GA with or without the company of notetaking. The difference regarding the levels of each GA type, as reported in Chapter 3, did not reach statistical significance. Similar data were obtained from the experiment to investigate whether the previous funding would sustain in a controlled environment.

4.2.3.1 Descriptive Data

The following two scatter plots (Figure 27) illustrate the counts of each GA type under the easy and difficult conditions. An overall observation[2] would immediately reveal that under both conditions, there are considerable individual differences among interpreters regarding the frequency of GA, echoing with the overall feature that GA is subject to individual differences. It is also evident that most interpreters were associated with both GA+NT and GA-NT in the experiment, which accords with the previous finding in Chapter 3. Specifically, as shown in the plots, 4 out of 7 interpreters conducted both GA types when processing the easy speech. By contrast, 3 interpreters, when processing the easy speech, did not interrupt producing notes while looking away from their notepads. Under the difficult speech, 8 out of 9 interpreters performed a mixture of GA+NT and GA-NT when the source speech was difficult. The 1 interpreter who only resorted to one type of GA completely detached from taking notes when averting gaze from the notepad.

[2] The number of interpreters who performed GA was 8 and 10 in easy and difficult speech, respectively. However, one of the Livescribe SmartPens did not successfully pick up the notes of one interpreter, and therefore the data regarding GA type was obtained from 7 and 9 interpreters in easy and difficult speech, respectively.

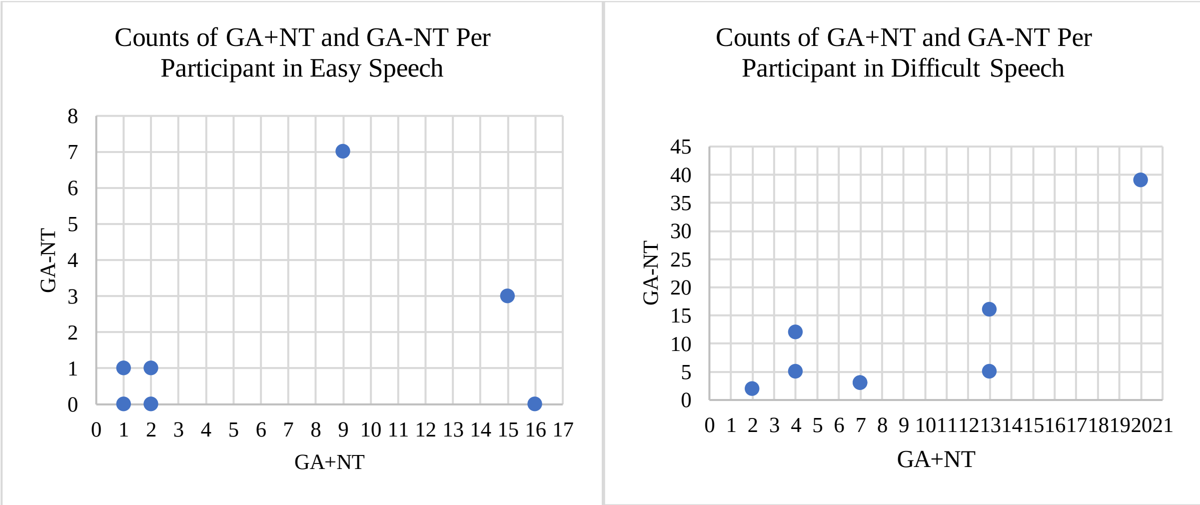
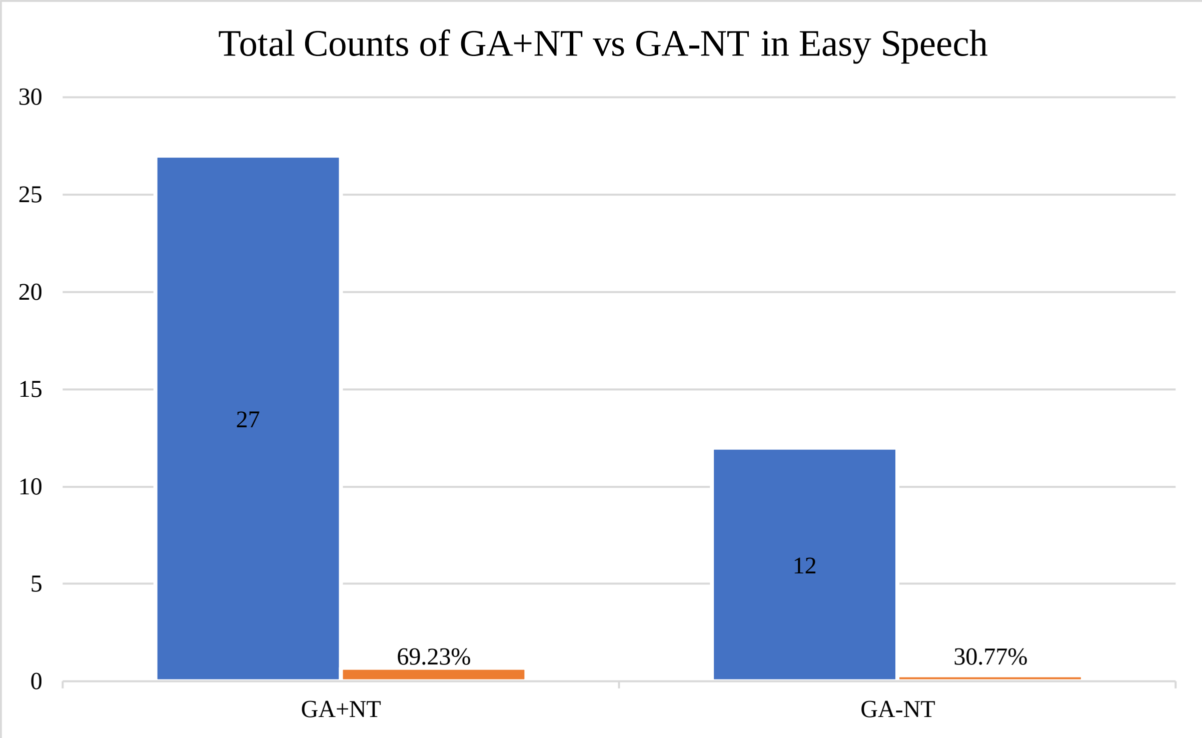


Figure 27 Counts of each GA type in each speech

On average, each interpreter performed 6.57 (*SD* = 6.71, *max* = 16, *min* = 1) and 7.78 (*SD* = 6.40, *max* = 20, *min* = 0) times of GA+NT under the easy and difficult speech, respectively. For GA-NT, the average count was 1.71 (*SD* = 2.56, *max* = 7, *min* = 0) per interpreter in the easy speech, and 9.89 (*SD* = 11.88, *max* = 39, *min* = 2) per interpreter under the difficult speech. In terms of the average length, each interpreter spent 0.70 (*SD* = 0.40, *max* = 1.44, *min* = 0.11) and 0.78 (*SD* = 0.42, *max* = 1.66, *min* = 0) seconds in performing GA+NT in easy and difficult speech, respectively.

The following bar charts were plotted to illustrate the distribution of each GA type in easy (Figure 28) and difficult (Figure 29) speech. To ensure the accuracy of analysis, data of the interpreters who only resorted to one type of GA is excluded. When processing the easy speech, 27 out of 39 GA was accompanied by notetaking, accounting for nearly 70% of the GA counts. By comparison, only 12 GA was performed while the action of taking notes was absent, contributing to a proportion of 30.77%. In terms of the total GA length, the relevant interpreters collectively spent 17.94 seconds (71.93%) averting their gaze while remaining engaged in notetaking. By comparison, a total of 7 seconds were contributed to averting gaze while interpreters were detached from notetaking, accounting for 28.07% of the total time spent on the GA-NT.



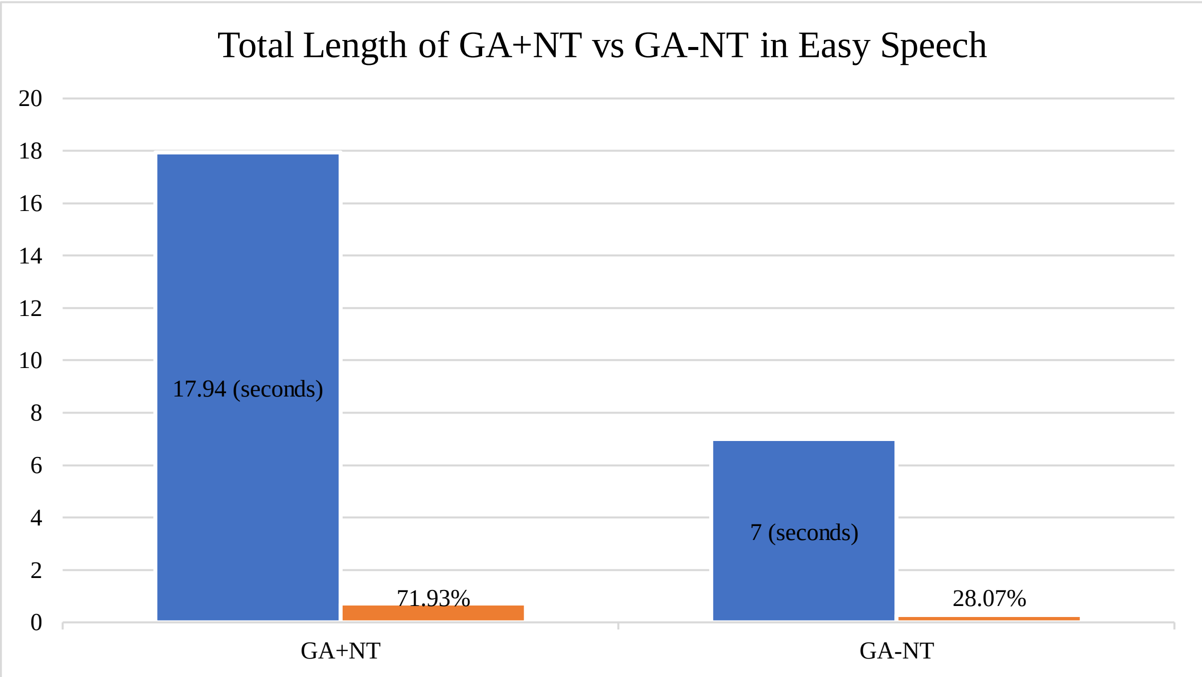


Figure 28 Distribution of each GA type under the easy speech

Interestingly, as shown in figure 29, under the difficult speech condition, the pattern where GA+NT was favoured over GA-NT has witnessed a reversion regarding both measurements. In specific, the interpreters who averted gaze when processing the difficult speech collectively conducted 85 times of GA-NT that amounted to a total length of 78.8 seconds, taking up 54.84% and 54.68% in the two measurements, respectively. By comparison, GA+NT was performed 70 times with a total length of 65.31 seconds, taking up 45.16% and 45.32% in the two measurements, respectively.

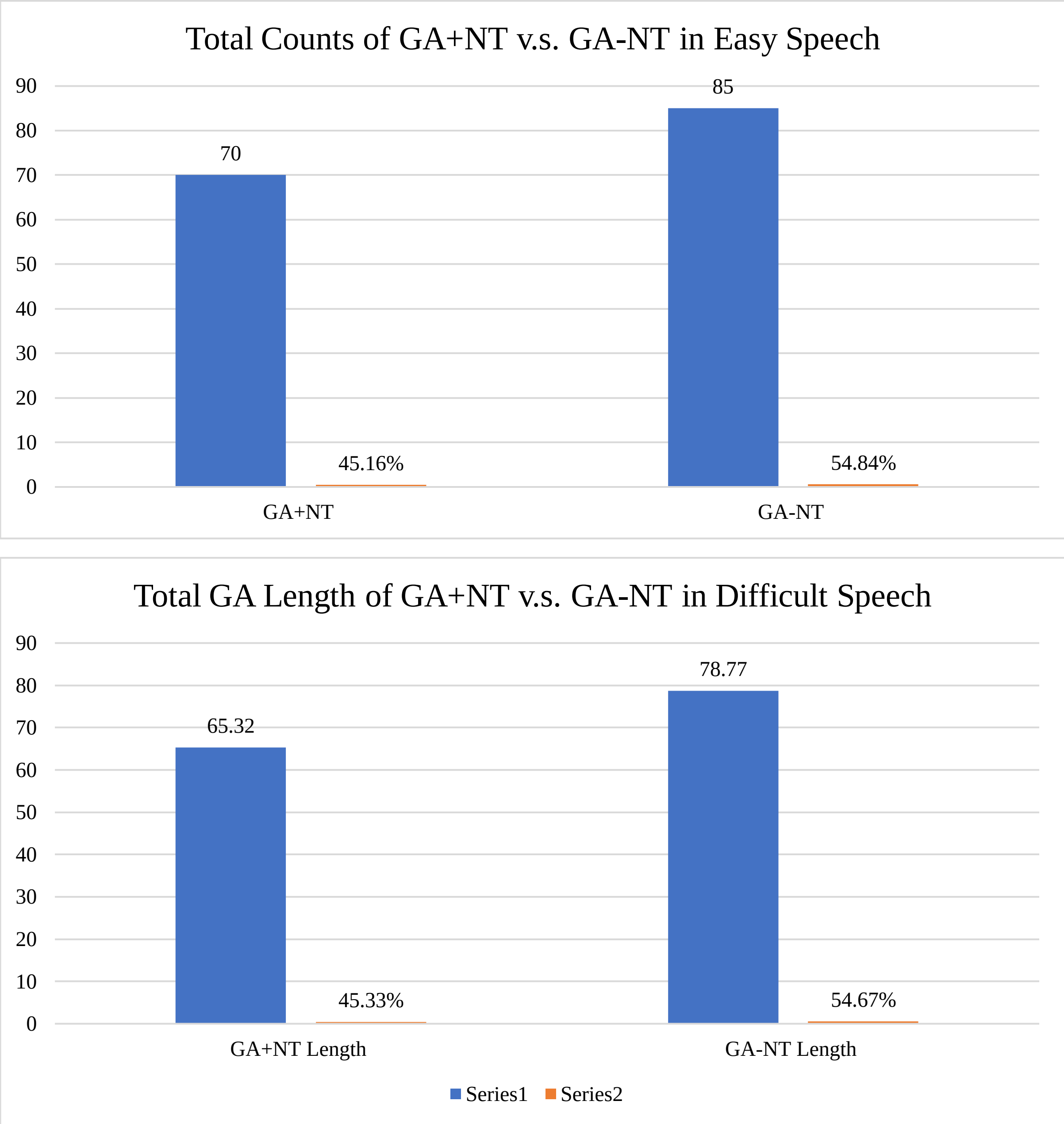


Figure 29 Distribution of each GA type under the difficult speech

Apart from the opposite preference observed under easy and difficult speech conditions, another difference between the two speech conditions was that the discrepancy between GA+NT and GA-NT in terms of counts and length was much more considerable when the speech difficulty was lower.

4.2.3.2 Inferential Analysis

4.2.3.2.1 Overall Comparison

Inferential analyses were conducted to investigate whether the above-mentioned differences between the two types of GA regarding counts and average GA length in each condition have reached statistical significance. The data from the interpreters who did not perform both types of GA was not included in the tests. To examine the assumption of normality, a Shapiro-Wilk test was conducted for all 8 data groups before selecting a robust means of analysis.

Table 15 Summary of Shapiro-Wilk test results

GA Type	***Shapiro-Wilk Test Result***	***Evidence for Departure from Normality***
GA+NT Counts (Easy Speech)	*w*(7) = .79, *p* = .03	Yes
GA-NT Counts (Easy Speech)	*w*(7) = .79, *p* = .01	Yes
GA+NT Average Length (Easy Speech)	*w*(7) = .93, *p* = .52	No
GA-NT Average Length (Easy Speech)	*w*(7) = .85, *p* = .13	No
GA+NT Counts (Difficult Speech)	*w*(7) = .94, *p* = .62	No
GA-NT Counts (Difficult Speech)	*w*(7) = .76, *p* = .02	Yes
GA+NT Average Length (Difficult Speech)	*w*(7) = .87, *p* = .20	No
GA-NT Average Length (Difficult Speech)	*w*(7) = .96, *p* = .78	No

As indicated in Table 15 above, the Shapiro-Wilk test found evidence for significant departure from normality for count measures for GA+NT and GA-NT under easy speech conditions. A significant departure from normality was also indicated by the test for counts of GA-NT under the difficult speech condition. Based on the normality check results, a parametric paired sample *t*-test and non-parametric Wilcoxon Signed Rank test were chosen accordingly.

In terms of counts, for GA coded under the easy speech, the Wilcoxon Signed Rank test revealed that GA+NT (*Md* = 2, *n* = 7) was significantly more frequent than GA-NT (*Md* = 1, *n* = 7), *z* = - 2.21, *p* = .027, with a large effect size, *r* = .59 (Cohen, 2013; Gignac and Szodorai, 2016). Contrarily, for GA occurred during the difficult speech, the Wilcoxon Signed Rank indicated that the difference between the counts of GA+NT (*Md* = 7, *n* = 9) and GA-NT (*Md* = 5, *n* = 9) did not reach statistical significance, *z* = - .49, *p* = .62. Regarding average length of GA under the easy speech condition, a paired sample t-test did not find statistically significant difference between GA+NT (*M* = .70, *Sd* = .40) and GA-NT (*M* = .29, *Sd* = .29), $t(6)= 1.83$, $p = .12$. For the average length of GA coded under the difficult speech, the *t*-test did not find a significant effect between GA+NT (*M* = .78, *Sd* = .42) and GA-NT (*M* = .98, *Sd* = .47), $t(8) = -1.36$, $p = .21$.

4.2.3.2.2 Comparing GA+NT and GA-NT in the Difficult Speech

As mentioned previously in 4.2.1.2.4, Speech B contains an equal portion of three types of sentences, namely syntactically complex sentences, syntactically simple sentences that contain LFWs, and syntactically complex sentences containing LFWs. The exploration of the potentially significant effects between GA+NT and GA-NT would not be thorough unless the differences regarding GA types among the three types of sentences were rigorously compared. Towards this goal, the overall data pertaining to GA types in the difficult speech was broken down into 12 data groups to be inferentially analysed.

Before a robust test was chosen, a Shapiro-Wilk test was conducted to test the assumption of normality for the scores of all 12 groups. Table 16 below summarises the results of the normality check. As shown in the table, the Shapiro-Wilk test did not find evidence for significant departure from normality for all measurements of average length. By contrast, the test indicated that the scores for most of the count measurements were not normally distributed. Based on the results of the Shapiro-Wilk test and a visual inspection of histograms and Q-Q plots, non-parametric and parametric tests were selected to compare the scores of count and length measurements, respectively. Specifically, for count measurements, a Wilcoxon Signed Rank test was chosen to test the within-group differences (such as counts of GA+NT vs GA-NT for syntactically complex sentences without LFWs), whereas a Mann Whitney test was deemed robust to test the between-group differences (i.e., GA+NT counts under different types of sentences). For length measurements, a paired sample t-test and an independent sample t-test were used for within and between-group comparisons, respectively.

Table 16 Summary of normality check results for count and length measurements for three types of sentences

GA Type	***Shapiro-Wilk Test Results***	***Evidence for Significant Departure from Normality***

GA+NT(CSS) Counts	*w*(9) = .87, *p* = .122	No
GA-NT (CSS) Counts	*w*(9) = .83, *p* = .046	Yes
GA+NT (CSS) Average Length	*w*(9) = .88, *p* = .15	No
GA-NT (CSS) Average Length	*w*(9) = .95, *p* = .65	No
GA+NT (LFW) Counts	*w*(9) = .72, *p* = .002	Yes
GA-NT (LFW) Counts	*w*(9) = .68, *p* < .001	Yes
GA+NT (LFW) Average Length	*w*(9) = .89, *p* = .21	No
GA-NT (LFW) Average Length	*w*(9) = .94, *p* = .61	No
GA+NT (Both) Counts	*w*(9) = .90, *p* = .25	No
GA-NT (Both) Counts	*w*(9) = .78, *p* = .01	Yes
GA+NT (Both) Average Length	*w*(9) = .97, *p* = .88	No
GA-NT (Both) Average Length	*w*(9) = .86, *p* = .09	No

In terms of count measurements, the Wilcoxon Signed Rank test compared the within-subjects differences between GA+NT and GA-NT under the three sentence types to investigate if one type of GA was significantly more frequent than the other. For syntactically complex sentences that did not contain rare words, the test did not find statistically significant difference between GA+NT (*Md* = 2, *n* = 9) and GA-NT (*Md* = 1, *n* = 9), *z* = -.07, *p* = .94. For syntactically simple sentences that contain LFWs, the test also failed to indicate significant differences between GA+NT (*Md* = 1, *n* = 9) and GA-NT (*Md* = 2, *n* = 9), *z* = -1.84, *p* = .07. For the sentences that contain both complex structures and rare words, a similar non-significant effect was rendered between GA+NT (*Md* = 2, *n* = 9) and GA-NT (*Md* = 1, *n* = 9), *z* = -1.41, *p* = .89.

A non-parametric Mann-Whitney test was conducted to examine the between-group difference in terms of counts for both GA+NT and GA-NT. For GA+NT, the Mann Whitney test did not find statistical significance between LFW-free complex sentences (*Md* = 2, *n* = 9) and syntactically simple sentences containing LFW (*Md* = 1, *n* = 9), *U* = 31.5, z = - .82, *p* = .41. The test also failed to render a significant effect between sentences containing both difficulty variables (*Md* = 2, *n* = 9) and complex sentences composed with common words (*Md* = 2, *n* = 9), *U* = 33.5, z = - .63, *p* = .53. For the last pair – simply structured sentences with LFWs (*Md* = 1, *n* = 9) and sentences with both variables (*Md* = 2, *n* = 9), the Mann Whitney Test did not find evidence for statistically significant difference, *U* = 22, z = - 1.68, *p* = .09.

For GA-NT, the Mann Whitney test did not render a significant effect between complex sentences composed without LFW (*Md* = 1, *n* = 9) and syntactically simple sentences containing LFW (*Md* = 2, *n* = 9), *U* = 29.5, z = - .984, *p* = .33. The comparison between sentences containing both difficulty variables (*Md* = 1, *n* = 9) and LFW-free complex sentences (*Md* = 1, *n* = 9), *U* = 40.5, z = .00, *p* = 1.00. Finally, between syntactically complex sentences containing rare words (*Md* = 1, *n* = 9) and syntactically simple sentences containing LFW (*Md* = 2, *n* = 9), the Mann Whitney Test also failed to reveal evidence for a significant effect, *U* = 31, z = -.85, *p* = 3.9.

In terms of length measurements, a paired sample *t*-test was conducted to compare the within-group differences between the length of GA+NT and GA-NT coded under three types of sentences to examine if the average length of one type of GA is significantly longer than the other. For syntactically complex sentences composed with simple words, the test did not find statistically significant difference between GA+NT (*M* = .58, *Sd* = .50) and GA-NT (*M* = .88, *Sd* = .23), *t*(8)= -1.13, *p* = .29. For syntactically simple sentences that contained LFWs, the *t*-test did not render a significant effect between GA+NT (*M* = .44, *Sd* = .37) and GA-NT (*M* = .86, *Sd* = .60), *t*(8)= -1.77, *p* = .12. Finally, for the sentences that contained both complex structures and rare words, the *t*-test also failed to render statistically significant differences between GA+NT (*M* = .89, *Sd* = .57) and GA-NT (*M* = .59, *Sd* = .49), *t*(8)= 1.02, *p* = .34. Echoing with the previous finding, the above inferential analyses confirmed the previous finding in Chapter 3 where the average lengths of GA+NT and GA-NT were not subject to statistical differences.

To test the between-group differences in terms of the average GA length, an independent sample *t*-test was performed to compare the average length of GA+NT and GA-NT across the three sentence types. For GA+NT, the *t*-test did not render a significant effect between syntactically complex sentences composed with common words (*M* = .58, *Sd* = .50) and syntactically simple sentences that contained LFWs (*M* = .44, *Sd* = .37), *t*(16) = .65, *p* = .53. The comparison between syntactically complex sentences composed with common words (*M*

= .58, *Sd* = .50) and sentences containing both variables (*M* = .89, *Sd* = .57), according to the *t*-test results, also failed to render a significant effect, $t(16) = -1.23$, $p = .238$. Finally, for the last pair - syntactically simple sentences that contained LFWs (*M* = .44, *Sd* = .37) and sentences containing both variables (*M* = .89, *Sd* = .57), the *t*-test did not reveal statistical difference, $t(16) = -1.96$, $p = .07$.

For GA-NT, the length scores were also compared across the three sentence types via an independent sample *t*-test. Between the LFW-free sentences that were syntactically complex (*M* = .88, *Sd* = .68) and simply structured sentences containing LFWs (*M* = .86, *Sd* = .60), the *t*-test did not reveal evidence for statistical significance, $t(16) = .06$, $p = .96$. The comparison between syntactically complex sentences composed with common words (*M* = .88, *Sd* = .68) and sentences containing both difficulty elements (*M* = .59, *Sd* = .49) also failed to indicate a significant effect, $t(16) = .1.02$, $p = .33$. Finally, the *t*-test did not render a significant effect between syntactically simple sentences containing rare words (*M* = .86, *Sd* = .60) and sentences containing both difficulty variables (*M* = .59, *Sd* = .49), $t(16) = .1.04$, $p = .32$.

4.2.3.3 Summary

In summary, GA was significantly more often accompanied by the action of notetaking when the speech was easy. In other conditions, the counts and length of GA+NT and GA-NT did not differ significantly.

4.2.4 Priming Effect

4.2.4.1 Descriptive Data Regarding GA

To recap, the second RQ asked whether priming subjects with LFWs would, to a certain degree, alleviate cognitive load and lead to less GA. The rationale primarily rests on the notion that exposure to LFWs could reduce the word frequency effect (Keuleers *et al.*, 2010; Monaghan *et al*., 2017; Stevens, 2017). To create a comparison, the time allowed for priming was 8 minutes for one chunk in the difficult speech and 4 minutes for another chunk. It was hypothesised that compared with the shorter priming condition, the longer exposure condition would see a reduced GA level both in frequency and dwelling time. Distribution of GA measures under the two conditions are illustrated in the following box plots (Figure 30) per the master sheets in Appendix J.

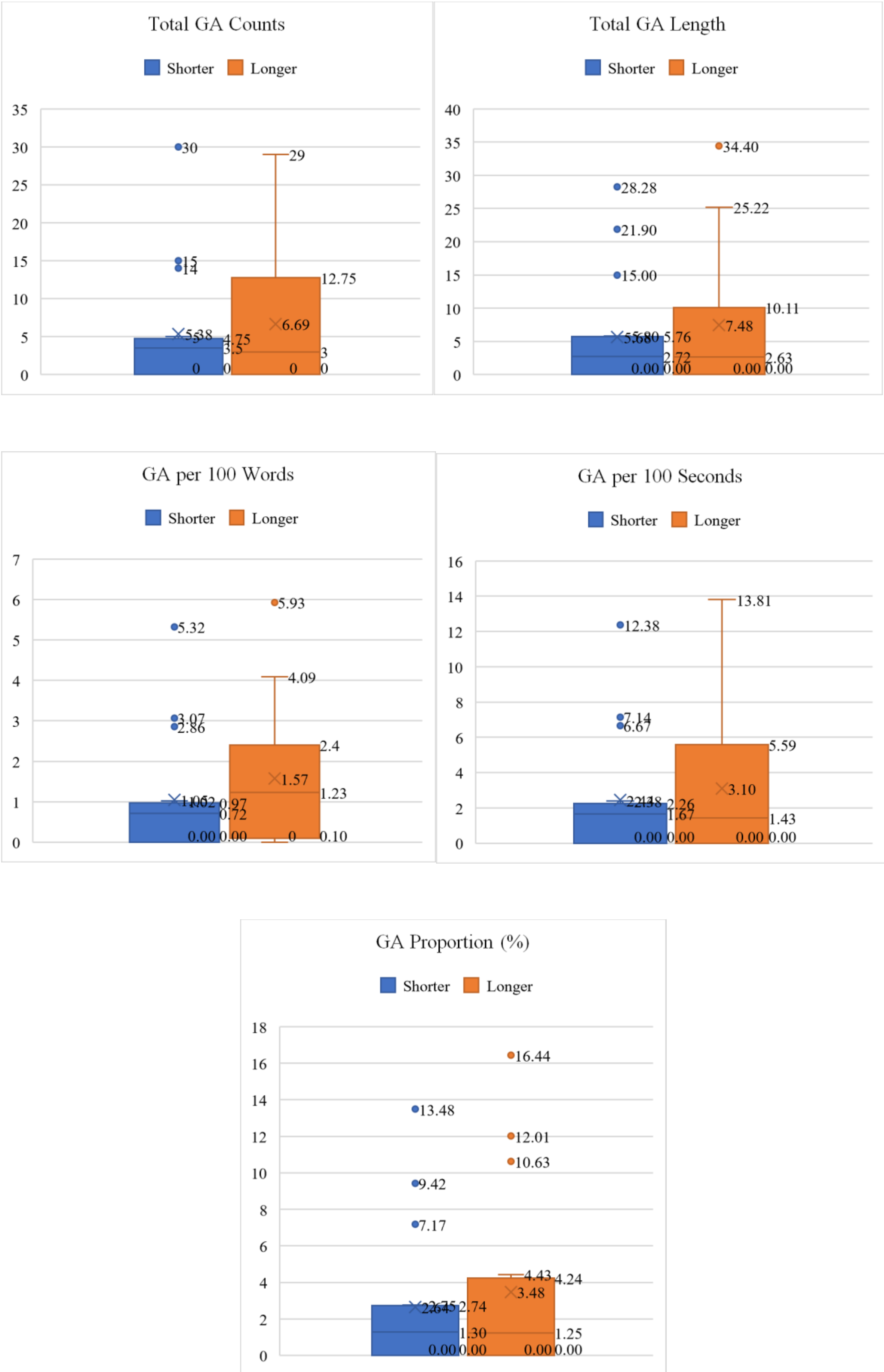


Figure 30 Box plots for GA measurements under longer and shorter exposure

Under the shorter exposure condition, the median score of GA counts is 3.5, covering an interquartile range between 0 (Q1) and 4.75 (Q3), meaning that half of the participants who underwent the shorter priming session performed less than 4.75 times of GA. On average, 5.38 times of GA were coded for each participant (*max* = 30, *min* = 0). By comparison, under the longer exposure condition, the median score of GA counts is 3, with an interquartile range between 0 (Q1) and 12.75 (Q3), suggesting that half of the interpreters who were primed longer performed up to nearly 13 times of GA. On average, 6.69 times of GA were coded under the longer exposure condition (*max* = 29, *min* = 0).

Regarding total GA length, under the shorter priming condition, the median score is 2.72 seconds, with an interquartile range between 0 (Q1) and 5.8 seconds (Q3), suggesting that half of the interpreters spent up to 5.8 seconds looking away from notepads. On average, each interpreter invested 5.68 seconds in GA (*max* = 34.4, *min* = 0). Under the longer exposure condition, the median score for total GA length is 7.48 seconds. The interquartile range is from 0 (Q1) to 10.11 seconds (Q3), suggesting that half of the interpreters spent up to 10.11 seconds averting gaze from notepads. On average, each participant spent 7.48 seconds in GA (*max* = 34.4, *min* = 0).

As for GA coded every 100 words of the source speech, under the shorter exposure condition, the median score is 0.72, with an interquartile range between 0 (Q1) and 0.97 (Q3). On average, each interpreter conducted 1.05 times of GA for every 100 words from the source speech (*max* = 5.32, *min* = 0). By contrast, the median score for the same measurement under the longer exposure condition is 1.23, with an interquartile range between 0.1 (Q1) and 2.4 (Q3), suggesting that 0.1 to 2.4 times of GA were coded for every 100 words uttered. On average, 1.57 times of GA were conducted for every 100 words of the source speech (*max* = 5.93, *min* = 0).

In terms of GA coded for every 100 seconds of the source speech, under the shorter priming session, the median score of GA counts is 1.67, covering an interquartile range between 0 (Q1) and 2.26 (Q3), meaning that half of the participants who underwent the shorter priming session performed less than 2.26 times of GA while processing per 100 seconds of the source speech. On average, 2.44 times of GA were coded for each 100 seconds (*max* = 12.38, *min* = 0). Compared with shorter priming, the median score for the same measurement under the longer exposure condition is 1.43, with an interquartile range between 0 (Q1) and 5.59 (Q3), suggesting that up to 5.59 times of GA were coded for each 100 seconds. On average, each participant under the longer exposure condition performed 3.1 times of GA for every 100 seconds of the speech (*max* = 13.81, *min* = 0)

Finally, regarding the ratio of time spent in GA to the total speech length, under the shorter exposure condition, the median score is 1.3% with an interquartile range between 0% (Q1) and 2.74% (Q3), meaning that for half of the interpreters under the condition, 5.75 seconds were spent conducting GA. On average, 2.64% of the speech time (or 5.54 seconds) was invested in GA (*max* = 13.48%, *min* = 0%). By contrast, under the longer priming condition, the median score for the same measurement is 1.25%, with an interquartile range between 0% (Q1) and 4.24% (Q3), meaning that for half of the interpreters under the condition, 8.9 seconds were spent conducting GA. On average, 3.48% of the speech time (7.31 seconds) was spent performing GA (*max* = 16.44%, *min* = 0%).

It is noticeable from the above boxplots that all data distributions contain a certain degree of skewness and that almost all measurements have outliers located outside the upper whiskers.

By observing the mean values from the above table, it would seem warranted to suggest that GA level under the longer priming condition was slightly higher than GA level under the shorter exposure condition. Specifically, as indicated by count-related measures, GA occurred more often under the longer exposure condition. Also, the time spent on GA, as indicated by the GA proportion (3.33% vs 2.64%), was longer in the longer priming condition. The descriptive data do not seem to support the hypothesis developed for the RQ. However, the inferential analysis would be needed to confirm whether the difference reached statistical significance.

4.2.4.2 Inferential Analysis for Priming and GA

A Shapiro-Wilk test was conducted to examine whether the scores in the above table were normally distributed. As shown in Table 17 below, the test found evidence for significant departure from normality for all data groups. Based on the normality check results and the need to compare data in pairs, a non-parametric Wilcoxon Signed Rank test was employed.

Table 17 Shapiro-Wilk Test results for GA measures under longer and shorter priming session

GA Type	***Shapiro-Wilk Test Result***	***Evidence for Departure from Normality***
GA Counts (Longer Exposure)	$w(16) = .79, p = .002$	Yes
GA Counts (Shorter Exposure)	$w(16) = .68, p <.001$	Yes
Avg. GA Length (Longer Exposure)	$w(16) = .86, p = .016$	Yes
Avg. GA Length (Shorter Exposure)	$w(16) = .88, p = .038$	Yes
GA/100 Words (Longer Exposure)	$w(16) = .80, p = .003$	Yes
GA/100 Words (Shorter Exposure)	$w(16) = .71, p <.001$	Yes
GA/100 Seconds (Longer Exposure)	$w(16) = .79, p = .002$	Yes
GA/100 Seconds (Shorter Exposure)	$w(16) = .71, p <.001$	Yes
GA Proportion (Longer Exposure)	$w(16) = .74, p <.001$	Yes
GA Proportion (Shorter Exposure)	$w(16) = .71, p <.001$	Yes

The Wilcoxon test failed to reveal any significant effect for all pairs. Specifically, GA counts under longer exposure condition ($Md = 3$, $n = 16$) and shorter exposure condition ($Md = 3.5$, $n = 16$) were not significantly different, $z = -1.19$, $p = .233$. GA per hundred words under longer exposure condition ($Md = 1.13$, $n = 16$) did not show any statistical difference from under shorter exposure condition ($Md = .72$, $n = 16$), $z = -1.45$, $p = .15$. In terms of GA per hundred seconds of source speech length, there was no significant effect between longer ($Md = 1.43$, $n = 16$) and shorter exposure ($Md = 1.67$, $n = 16$), $z = -1.38$, $p = .17$. For length measures, there was no effect between the average GA length under longer priming ($Md = .72$, $n = 16$) and shorter priming conditions ($Md = .72$, $n = 16$), $z = -.051$, $p = .96$. Finally, in terms of the time spent in averting GA, or GA proportion, the Wilcoxon test did not render significant difference between longer ($Md = 1.25$, $n = 16$) and shorter ($Md = 1.30$, $n = 16$) priming, $z = -1.17$, $p = .24$. Based on the inferential analysis results, priming did not prove effective in affecting GA levels.

4.2.4.3 Descriptive Data Regarding Priming and LFW Rendition

Although priming was adopted to primarily test whether exposing interpreters to upcoming LFWs would lead to a reduction of WFE, and thereby reducing cognitive costs associated with processing the difficult speech, the effect could arguably also manifest itself in faster lexical retrievals as it has been well documented that exposure plays a critical role in reducing WFE (Keuleers *et al*., 2010; Monaghan *et al*., 2017; Stevens, 2017). To test if priming would affect lexical retrieval, hit rates (the ratio of LFWs that were faithfully rendered) obtained under each chunk of difficult source speech were compared.

A box plot (Figure 31) below is constructed to display the distribution of the average hit rates under shorter and longer exposure conditions for all interpreters based on a table in Appendix K.

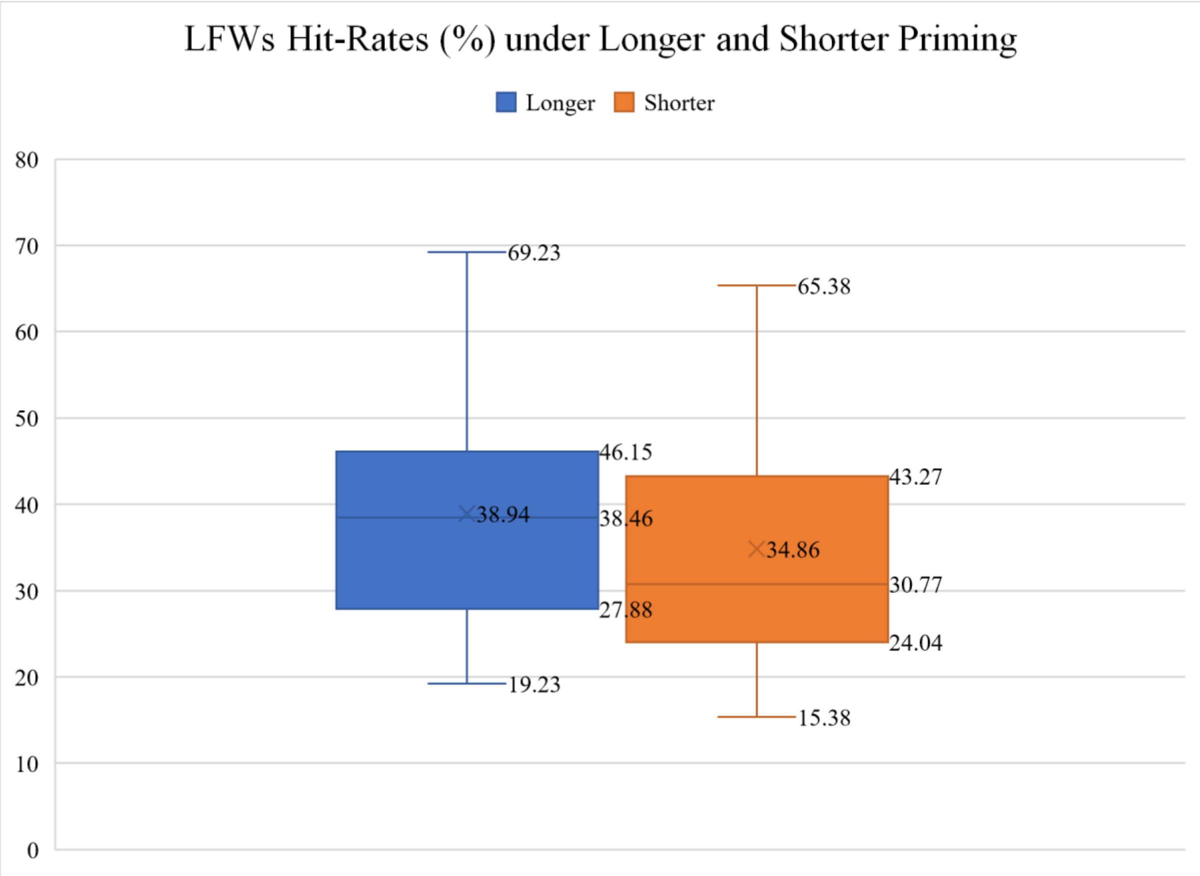


Figure 31 LFWs hit-rates under longer and shorter priming

As indicated in the box plot, the median score of the hit rate under the longer exposure condition is 38.46%, with an interquartile rang between 27.88% (Q1) to 46.15% (Q3), indicating that half of the participants who experienced the longer priming session had a hit rate between 27.88 % and 46.15%. On average, 38.94% of the LFWs were faithfully rendered (*max* = 69.23%, *min* = 19.23%). By contrast, the median score of the same measurement under the shorter exposure condition is 30.77%. The interquartile range is between 24.04% (Q1) and 43.27% (Q3), a range similar to the shorter exposure condition. On average, 34.86% of the LFWs were accurately interpreted (*max* = 65.38%, *min* = 15.38%).

4.2.4.4 Inferential Analysis Regarding Priming and LFW Rendition

Before the inferential analysis, A Shapiro-Wilk test was performed and showed that the hit rates under shorter exposure departed significantly from normality ($w(16) = .83$, $p = .009$). By contrast, the Shapiro-Wilk test did not find evidence of non-normality ($w(16) = .95$, $p = .52$). Based on the normality check results, a Wilcoxon test was performed to compare the difference between the two groups of data. The Wilcoxon test did not find any significant effect between the hit rates under longer priming period ($Md = 38.46$, $n = 16$) and shorter priming period ($Md = 30.77$, $n = 16$), $z = -1.67$, $p = .95$, suggesting that longer exposure time did not lead to more accurate rendition of the LFWs. In other words, there was no priming effect regarding lexical decisions.

4.2.4.5 Summary

Interpreters went through longer and shorter priming sessions where they were exposed to the full list of LFWs that were about to appear in the source speech. Priming effects were expected to manifest in reduced GA levels and higher hit rates. Unexpectedly, however, the inferential analysis did not find evidence for statistical significance in either direction. In other words, repetition priming, within the scope of this experimental study, did not lead to any priming effect.

4.2.5 GA and Interpreters' Performance

The third RQ pursued in this experimental study concerns the suggestion that GA boosts cognitive performance by allowing subjects to disengage from distractions, thereby channelling more cognitive resources to the task at hand (Glenberg *et al*., 1998). Within the scope of this experimental study, the hypothesis under RQ3 encompasses two aspects: GA would not boost interpreters' performance when the speech is easy because the cognitive load induced from processing the easy speech would be manageable; contrarily, under the difficult speech condition, GA levels would positively correlate with interpreters' performance scores because GA would help interpreters to channel more resources to processing the source speech, contributing to more thorough and accurate comprehension. The following sections are dedicated to reporting results obtained via descriptive and inferential analyses.

4.2.5.1 Descriptive Analysis

As reported in 4.2.1, GA occurred in both easy and difficult speech. To ensure accuracy of analysis, data obtained under the two difficulty conditions were collected and analysed separately. A box plot (Figure 32) below is constructed to illustrate the distribution of the performance scores from GA interpreters and non-GA interpreters when the source speech was easy.

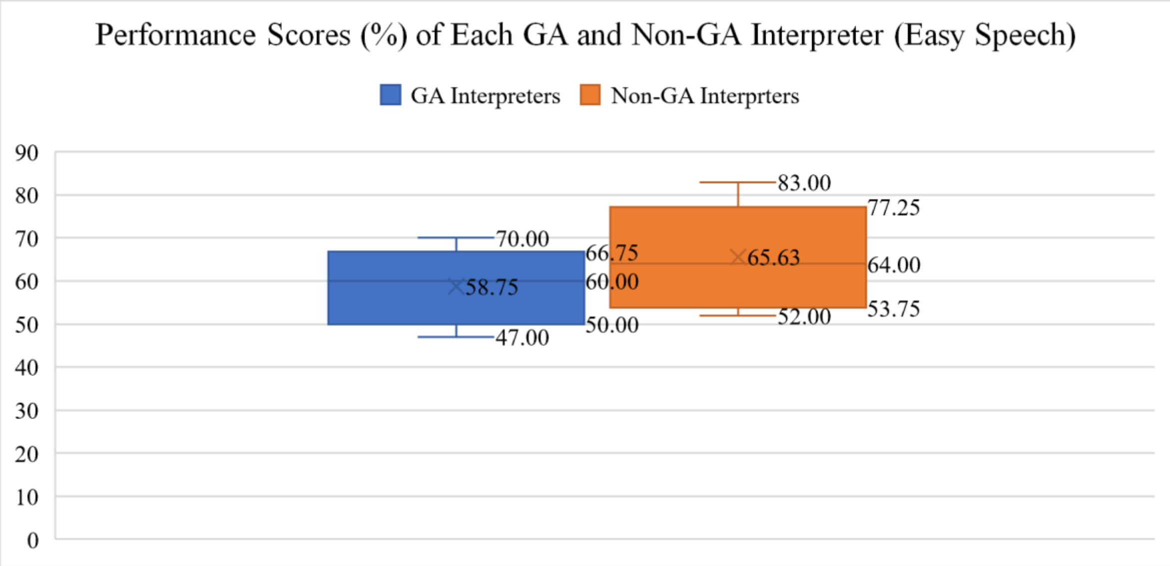


Figure 32 Performance scores of each GA and non-GA interpreters in the easy speech

The box plot above illustrates the distribution of performance scores (%) of each GA interpreter (those who averted gaze) and non-GA interpreter (those who did not avert gaze) for the easy speech. For GA interpreters, the median is 60, with an interquartile range between 50 (Q1) and 66.75 (Q3), suggesting that half of the GA interpreters achieved a score between 50 and 66.75. On average, each GA interpreter scored 58.75 (*max* = 70, *min* =47). In comparison, the box for non-GA interpreters shows a higher median (64). The interquartile range is between 53.75 (Q1) and 77.25 (Q3), meaning that half of the interpreters who did not avert gaze from notepads scored 53.75 and 77.25, higher than their counterparts' scores. On average, non-GA interpreters outperformed their counterparts with a score of 65.63 (*max* = 83, *min* = 52).

To gain a complete picture of how GA might affect interpreters' performance, within-group data was also obtained. In specific, among the interpreters who performed GA, the scores of GA sentences and non-GA sentences were calculated. The Figure 33 below illustrates the distribution of data.

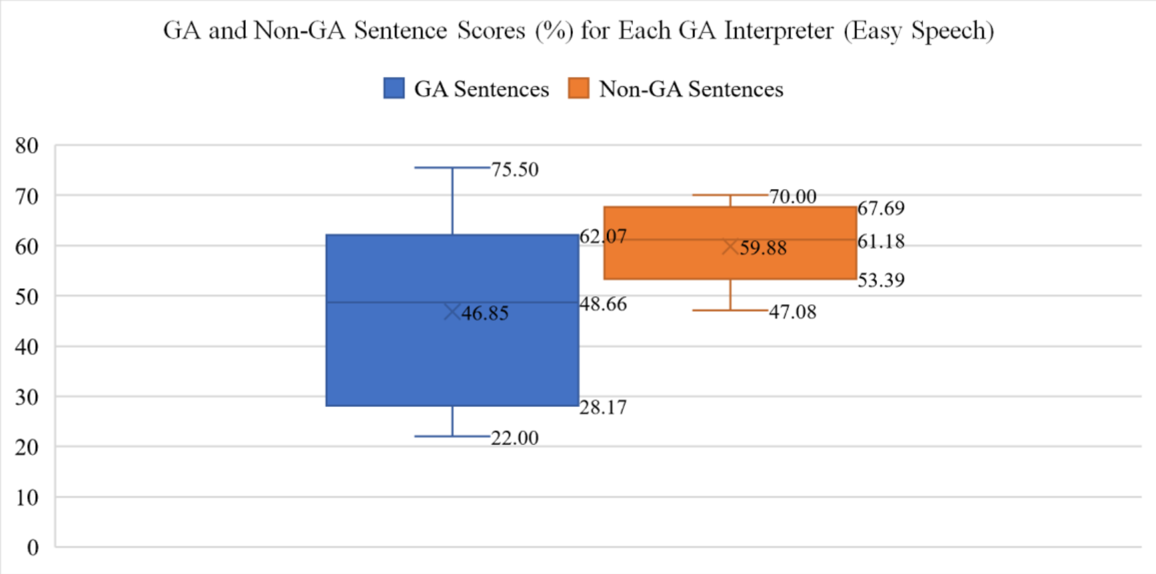


Figure 33 GA and non-GA sentence scores (%) for each GA interpreter in the easy speech

Figure 33 above compares the distribution of scores (%) of GA sentences (sentences where GA occurred) and non-GA sentences (sentences that are not associated with GA) for each GA interpreter under the easy speech condition. For GA sentences, the median is 48.66, with an interquartile range between 28.17 (Q1) and 62.07 (Q3), suggesting that half of the GA sentences were scored within the range. On average, each non-GA sentence was scored 46.85 (*max* = 75.50, *min* = 22). In comparison, the box for non-GA sentences shows a higher median (61.68). The interquartile range is between 53.39 (Q1) and 67.69 (Q3), meaning that half of the non-GA sentences were scored between the two figures. The average score for non-GA sentences is 59.88 (*max* = 70, *min* = 47.08).

As previously reported, as the speech difficulty increased, more interpreters conducted GA. The Cognitive Load Hypothesis (Glenberg *et al.*, 1998) and scholars (e.g., Doherty-Sneddon *et al.*, 2002; Phelps *et al.*, 2006) on the subject of GA have highlighted that GA emerges in cognitively challenging tasks and that the behaviour can boost concentration and accuracy of comprehension. Therefore, interpreters' renditions of the difficult source speech have significant bearings in understanding the cognitive value of GA. Figure 34 below illustrates the data distribution of performance scores obtained from GA and non-GA interpreters when the source speech imposed a significantly higher cognitive price.

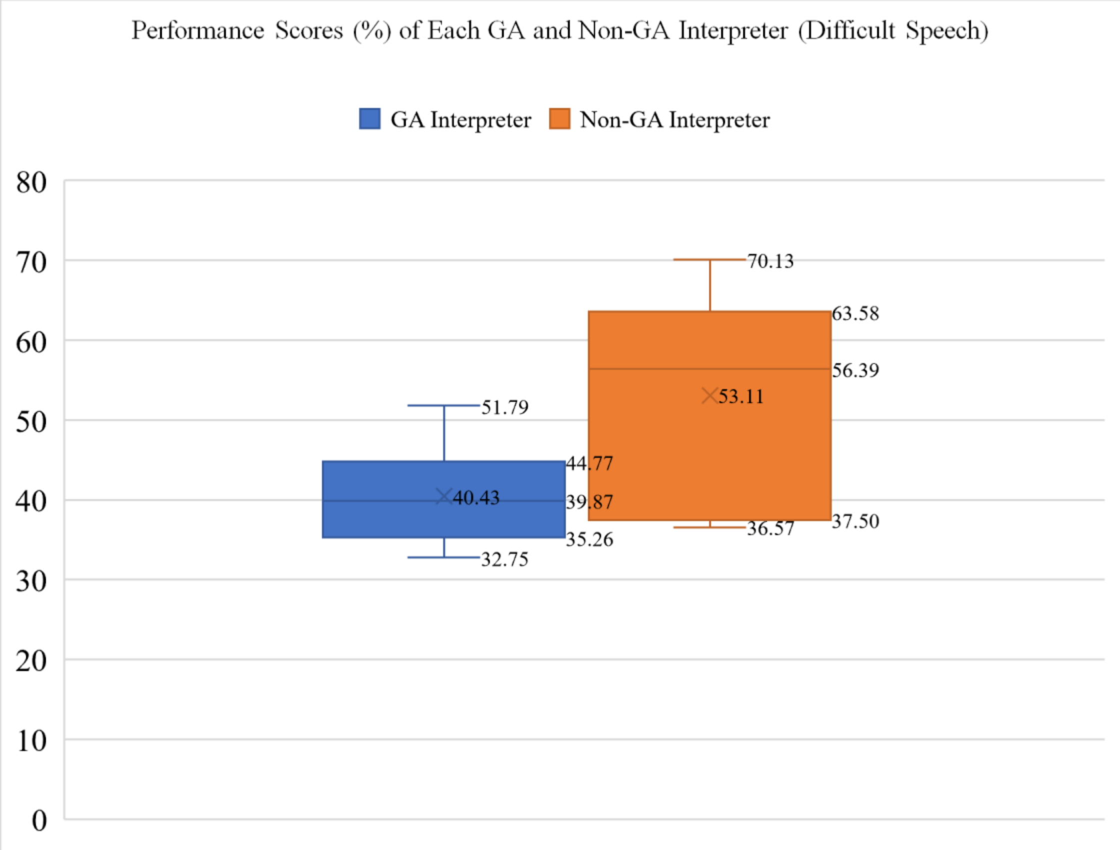


Figure 34 Performance scores of each GA and non-GA interpreters in the difficult speech

The above figure illustrates the distribution of performance scores (%) of each GA interpreter and non-GA interpreter for the difficult speech. For GA interpreters, the median is 39.87, with an interquartile range between 35.26 (Q1) and 44.77 (Q3), suggesting that half of the GA interpreters achieved a score within the range. On average, each GA interpreter scored 40.43 (*max* = 51.79, *min* = 32.75). In comparison, the box for non-GA interpreters shows a higher median (56.39). The interquartile range is between 37.50 (Q1) and 63.68 (Q3), meaning that half of the interpreters who did not avert gaze from notepads achieved a score within this range. On average, non-GA interpreters outperformed their counterparts with a score of 53.11 (*max* = 70.13, *min* = 36.57).

By observing the shapes of boxes and comparing the statistics, it would seem that the performance was better when GA did not occur. However inferential analyses are required to test if the differences are statistically significant. From the perspective of measuring cognitive load by task outcome, the decrease in average performance would suggest an increase in mental efforts (Sweller and Chandler, 1994; Sweller, 2011; Schroeder and Cenkci, 2018).

Consistent with the efforts invested in comparing the within-group differences when the source speech was easy, the performance score obtained for GA sentences and non-GA sentences were also compared. See Figure 35 below.

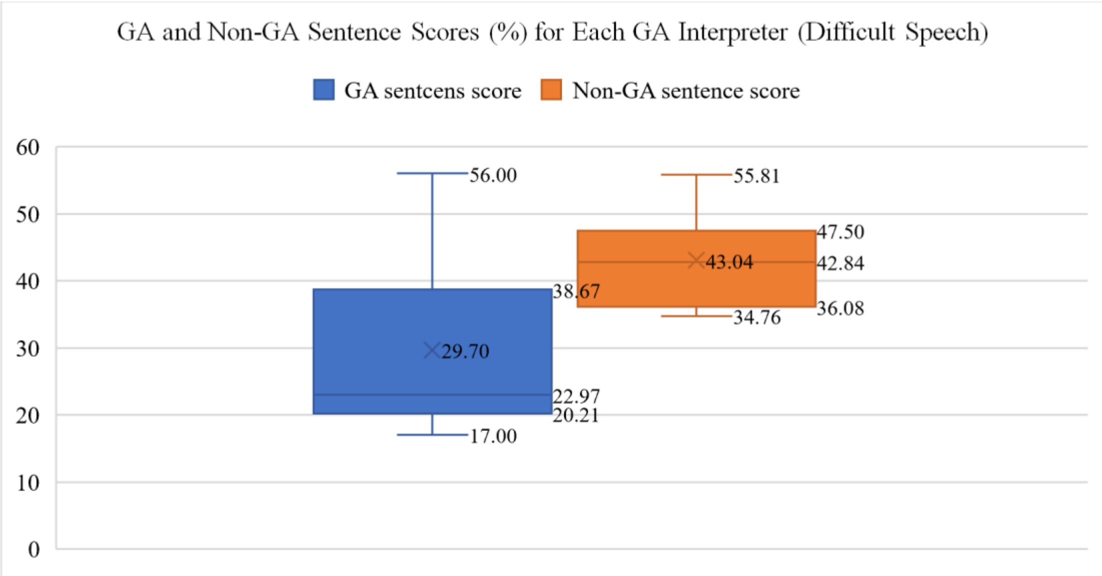


Figure 35 GA and non-GA sentence scores (%) for each GA interpreter in the difficult speech

The above box plot compares the distribution of scores (%) of GA sentences and non-GA sentences for each GA interpreter under the difficult speech condition. For GA sentences, the median is 22.97, with an interquartile range between 20.21 (Q1) and 38.67 (Q3), suggesting that half of the GA sentences were scored within the range. On average, each non-GA sentence was scored 29.7 (*max* = 56, *min* = 17). In comparison, the box for non-GA sentences shows a higher median (42.84). The interquartile range is between 36.08 (Q1) and 47.5 (Q3), a higher range than it of GA sentences. The average score for non-GA sentences is 43.04 (*max* = 55.81, *min* = 34.76).

In a nutshell, descriptive data comparing performance between GA and non-GA interpreters and GA and non-GA sentences have indicated that GA, instead of boosting performance, seemed to be associated with inferior performance. To test if the differences heretofore reported bear statistical significance, inferential analyses were conducted. The following section is dedicated to reporting the inferential results.

4.2.5.2 Inferential Analysis

As indicated in the above section, the scope of investigating if averting gaze during CI active listening and notetaking boosts performance encompassed the scoring of renditions not only by GA and non-GA interpreters but also of GA and non-GA sentences. Prior to selecting a test that would robustly examine if there existed the performance-enhancing function of GA in the context of CI, the assumption of normality would need to be tested, towards which a Shapiro-Wilk test was performed for the scores reported in the above tables. The normality check did not find evidence for significant departure from normality for all data groups. The test results are summarised in Table 18 as follows:

Table 18 Shapiro-Wilk test results for all Score groups

Data Group	***Shapiro-Wilk Test Result***	***Evidence for Departure from Normality***
GA Interpreters (Easy Speech)	$w(8) = .94$, $p = .64$	No
Non-GA Interpreters (Easy Speech)	$w(8) = .92$, $p = .42$	No
GA Sentences (Easy Speech)	$w(8) = .93$, $p = .57$	No
Non-GA Sentences (Easy Speech)	$w(8) = .95$, $p = .70$	No
GA Interpreters (Difficult Speech)	$w(6) = .98$, $p = .96$	No
Non-GA Interpreters (Difficult Speech)	$w(6) = .91$, $p = .44$	No
GA Sentences (Difficult Speech)	$w(6) = .83$, $p = .11$	No
Non-GA Sentences (Difficult Speech)	$w(6) = .88$, $p = .29$	No

As indicated by the *p* values, the Shapiro-Wilk did not find evidence for significant departure from normality. After a visual inspection of histograms and q-q plots, a parametric independent sample *t*-test was deemed methodologically sound to examine the differences between two comparable groups.

Under the easy speech condition, regarding the performance of GA interpreters (participants who averted gaze) and non-GA interpreters (participants did not perform GA), the *t*-test revealed that there was no significant statistical difference between the scores of GA interpreters ($M = 58.45$, $Sd = 8.42$) and non-GA interpreters ($M = 65.83$, $Sd = 11.61$), $t(14) = -1.46$, $p = .17$. By contrast, the test rendered that there was a significant difference between the scores of GA sentences (M = 46.85, Sd = 19.18) and non-GA sentences (M = 59.88, Sd = 8.03), $t(14) = -1.7$, $p = .04$, with large effect size, d = 0.87 (Cohen, 1988; Gignac and Szodorai, 2016), suggesting that non-GA sentences were rendered significantly better than GA sentences under the easy speech condition. Under the difficult speech, the independent *t*-test found significant effects between the performance scores of GA interpreters (M = 40.43, Sd = 6.17) and non-GA interpreters (M = 53.11, Sd = 13.48), $t(14) = -2.6$, $p = .021$, with a large effect size, d = 1.21 (Cohen, 1988; Gignac and Szodorai, 2016), suggesting that interpreters who did not avert their gaze while processing the difficult speech performed better than GA interpreters. Similarly, the *t*-test revealed statistically significant difference between the rendition performance of GA sentences (M = 29.70, Sd = 13.57) and non-GA sentences (M = 43.04, Sd = 7.14), $t(18) = -2.75$, $p = .001$, with a large size effect, d = 1.05 (Cohen, 1988; Gignac and Szodorai, 2016), indicating that GA sentences were rendered significantly worse than non-GA sentences when the speech was difficult.

4.2.5.3 Summary

To sum up, the data analysis did not find the performance-enhancing benefits suggested in the Cognitive Load Hypothesis (Glenberg *et al*., 1998). Instead of boosting performances, scores for sentences where GA was coded were significantly lower than the scores of non-GA sentences regardless of speech difficulty. Under the difficult speech condition, interpreters who did not resort to GA significantly outperformed their counterparts. The hypothesis developed for RQ3, therefore, is rejected.

4.2.6 Summary of key findings

For RQ1, it was found that interpreters did resort to GA during CI active listening and notetaking, verifying the corresponding finding from the corpus study. Similar to the corresponding finding in the corpus study, interpreters in the experimental study also demonstrated considerable individual differences. However, compared with the corpus study, where all participants were found to resort to GA, some participants in the experimental study did not perform GA at all.

As expected, the difficult speech saw significantly higher GA levels across all measures than the easy speech. The corresponding finding from the corpus study is, therefore, verified. Within the difficult speech, where interpreters' cognitive load was kept at a higher level, there was no significant effect among the three types of sentences in terms of GA frequency.

However, average GA length was significantly longer when syntactically complex sentences containing LFWs were experienced.

Also, under the umbrella of RQ1, GA+NT was only significantly more frequently than GA-NT when the speech was easy. Under the difficult speech, the difference between the two types of GA in terms of frequency did not reach statistical significance. For both types of the GA, average length, regardless of speech difficulty, did not vary significantly. The finding would also substantially verify the corresponding finding yielded from the corpus study.

For RQ2, the inferential analysis did not reveal a priming effect either in terms of accuracy of rendering the LFWs or GA levels, implying that priming interpreters with upcoming LFWs did not result in a reduced level of cognitive load. The hypothesis developed for RQ2 is, therefore, rejected.

For RQ3, the inferential analysis results failed to confirm the performance-enhancing function of GA. Surprisingly, regardless of speech difficulties, non-GA sentences were performed significantly better than GA sentences. Under the difficult speech condition, GA interpreters were significantly outperformed by non-GA interpreters. The results for RQ3 indicates that in the context of this study, GA, instead of boosting performance, was associated with performance deterioration.

4.3 Discussion

4.3.1 Overview

So far, this experimental study has been geared toward answering the following research questions:

1. *Would the findings yielded from Chapter 3 sustain in the experimental study?*
2. *Would priming interpreters with LFWs in the upcoming speech result in a reduced level of GA?*
3. *Would GA enhance interpreters' performance?*

Evidence was gathered and rigorously analysed to answer the above questions. The main findings are:

1. Answering RQ1- it was found that interpreters did resort to GA during CI active listening and notetaking. Some participants in the experimental study did not perform GA at all. Also, the difficult speech saw significantly higher GA levels across all measures than the easy speech. Within the difficult speech, where interpreters' cognitive load was kept at a higher level, there was no significant effect among the three types of sentences in terms of GA frequency. However, average GA length was significantly longer when syntactically complex sentences containing LFWs were experienced. Additionally, GA+NT was only significantly more frequently than GA-NT when the speech was easy. Under the difficult speech, the difference between the two types of GA in terms of frequency did not reach statistical significance. For both types of the GA, average length, regardless of speech difficulty, did not vary significantly.
2. Answering RQ2 - inferential analysis did not reveal a priming effect either in terms of accuracy of rendering the LFWs or GA levels, implying that priming interpreters with upcoming LFWs did not result in a reduced level of cognitive load.
3. Answering RQ3 - The inferential analysis results failed to confirm the performance-enhancing function of GA. Regardless of speech difficulties, GA and non-GA interpreters did not differ in performance scores. Surprisingly, however, a significant effect was found within the group of GA interpreters: GA sentences were performed significantly worse than non-GA sentences, indicating that in the context of this study, GA, instead of boosting performance, was associated with performance deterioration.

Similar to the corpus study, some findings above concern overarching aspects of understanding GA in CI active listening and notetaking. The General Discussion chapter (Chapter 5) is dedicated to interpreting the overarching findings from this study. To avoid repetition, the following section is dedicated to interpreting results germane to the experimental study.

4.3.2 Speech Difficulty and GA

4.3.2.1 Low Difficulty vs High Difficulty

As shown in 4.2.2.1, GA level was significantly higher under the difficult speech than under the easy speech. Specifically, GA was significantly more frequent and lasted significantly longer when the participants processed the difficult speech. From the cognitive load perspective, the result is expected because it would be irrefutable to suggest that the difficult speech imposed a much higher cognitive load on interpreters than the easy speech. As mentioned in 4.1.2.1.3.1, owing to the manipulation of independent variables – complex sentence structures and LFWs – the speech difficulty for the second speech was intentionally kept at a high level. As mentioned in 2.6.2.1, CI active listening requires unmitigated attention from interpreters because it largely relies on bottom-up processing, an approach that requires a listener to construct meanings of the speech in a gradual and fixed order starting from the phoneme level to the final discourse level (Vandergrift, 2004). According to Cutler and Clifton (1999), bottom-up processing propels listeners to adopt linguistic and non-linguistic knowledge incrementally and comprehensively. Therefore, the intrinsic cognitive load stemming from active listening alone would impose a high cognitive price.

CI notetaking, the visible and extra effort that accompanies active listening, would also be difficult under the difficult speech for at least two reasons. First, CI notetaking is a process of decision-making (Piolat *et al.*, 2005; Gile, 2009). It would be warranted to argue that processing the difficult speech would delay the decision-making process simply because the interpreter would need more time to comprehend the source speech. Second, considering the three basic principles of CI notetaking, that is, economy, instantaneous seizability, and individuality (Albl-Mikasa, 2017), taking notes from the difficult speech would have violated some, if not all, principles. It would be warranted to argue that interpreters find noting down a rare word difficult. The low word frequency would likely mean interpreters do not have an available symbol for the word. If the word carries an important semantic or syntactic role, provided that the interpreter has registered the word in his/her memory store, the interpreter will likely come up with a solution on the spot by either inventing a symbol or writing down the words in part or whole. Since CI notes are also supposed to indicate the logic and a sense of the source speech (Seleskovitch, 1978; Albl-Mikasa, 2017), processing complex sentence structure would naturally challenge the aspiration to organise notes in a neat and sense-manifesting fashion. In other words, cognitive load stemmed from notetaking was also high under difficult speech condition.

Figure 36 compares mark-ups under a sentence from the easy speech (left) and the difficult speech (right). The source sentences are transcribed in example 1 below.

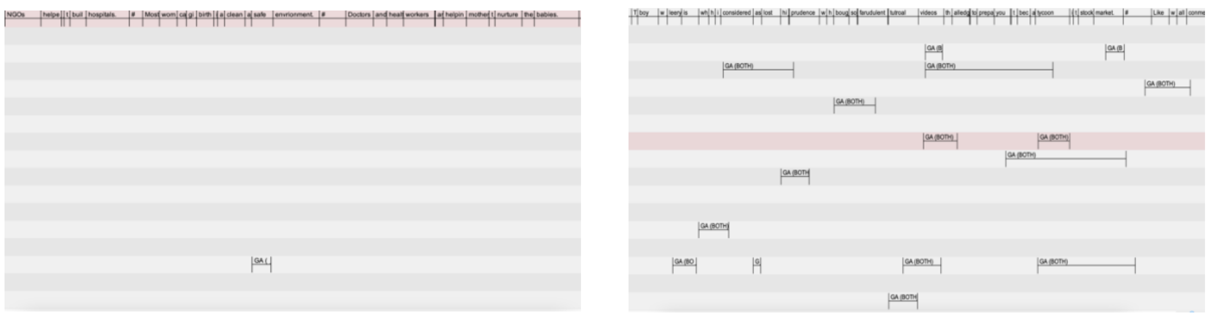


Figure 36 GA from two sentences (Sentences on the left attracted only one GA, whereas the sentence on the right is associated with a cluster of GA)

Example 1

Easy Sentence: *Most women can give birth in a clean and safe environment.*

Difficult Sentence: *The boy whom leery is what he is considered lost to his prudence when he bought some fraudulent tutorial videos that alleged to prepare you to become a tycoon in the stock market.*

As shown in the example, the easy sentence was constructed in a syntactically simple structure, whereas the difficult sentence contained LFWs ("*leery*", "*prudence*", and "*allege*") was composed in a syntactically complex manner by including several subordinate clauses, one of which being an object-relative clause ("*whom leery is what he is considered*"). A myriad of studies has shown that processing rare words entails a more cognitively effortful lexical retrieval process (e.g., Forster, 1976, 1981; Keuleers *et al.*, 2010; Vitevitch *et al.*, 2018; Popov and Reder, 2020) as memory search would consume more cognitive resources. Also, processing complex syntax can be cognitively expensive. For example, processing the object-relative clause would require interpreters to hold the clause in memory while waiting for the verb ("*lost*") to be uttered. In short, the difficult sentence would have imposed a much higher cognitive load than the easy sentence and therefore triggered a cohort of GA.

4.3.2.2 GA within the Difficult Speech

Having established that higher speech difficulty leads to higher GA levels, and considering the additive nature of cognitive load (Paas *et al*., 2003, 2005; Galy *et al*., 2012), it would be warranted to wonder if the source sentences that combined complex sentence structure and rare words, compared with the sentences that only contain one of the two variables, would be the most powerful in triggering GA. Inferential analysis results suggested that while there was no difference between the three types of sentences in terms of GA counts, the average length of GA was significantly longer when the sentences contained two variables than one. The pioneers in the research of GA have suggested that GA tends to occur at critical moments, and it allows individuals to shut down from distractions so that more resources can be re-routed to the task at hand (Glenberg *et al*., 1998; Doherty-Sneddon *et al*., 2002; Doherty-Sneddon and Phelps, 2005; Phelps *et al*., 2006). To reverse-engineer the mechanism of GA, it would be warranted to speculate that longer dwelling time in GA would signal the need for a higher volume of cognitive resources, resulting from higher cognitive load. Considering that cognitive load is additive (Paas *et al*., 2003, 2005; Galy *et al*., 2012), it would be reasonable for interpreters to engage in longer GA when they process source sentences that contain LFWs and are syntactically complex. In other words, interpreters dwelled longer on GA because they needed more resources to re-route. See the following example.

Example 2

Source Sentence: *However, my mother whom the harshness of life has* ***inured****, watched the video with me and* ***eulogised*** *the show as what she believes is that it is paramount to remind people that our time is defined by* ***treasury*** *which is unseen by any other times*

As shown in the example, the sentence contained several subordinate clauses and 3 LFWs (in bold). Based on the mark-ups in ELAN, this sentence attracted a 6-second-long GA (underlined). It would be warranted to argue that as the interpreter incrementally processed the sentence (Tanenhaus *et al*., 1995; Pickering and Gambi, 2018), her cognitive load would also have climbed rapidly due to the need to process the complex syntax. When LFWs began to be added to the speech, she would need to conduct a memory search, which would cause a sudden spike in processing load. As the load continued to accumulate, a deficit of cognitive resources would have occurred. To combat the overload, the interpreter would need to release a substantial amount of memory resources by performing a long GA.

4.3.3 GA+NT vs GA-NT

4.3.3.1 The Popularity of GA+NT in easy speech

An overall finding regarding GA types from the experimental study showed that most GA interpreters performed a mixture of GA+NT and GA-NT. Descriptive data suggested that when the source speech was easy, it was more frequent for interpreters to avert their gaze while maintaining notetaking than detaching from notetaking. However, under the difficult speech, GA-NT was found to be more prevalent. The inferential analyses found that the differences regarding GA counts and length were not statistically significant across all conditions with only exception: GA+NT was significantly more frequent than GA-NT ($p = .027$) under the easy speech condition.

To briefly recap, as previously discussed in chapter 3, it is reasonable for interpreters to stay engaged in notetaking during GA. By taking notes, interpreters not only take advantage of the encoding and storage function of notetaking (DiVesta and Gary, 1972; Kiewra, 1985; Albl-Mikasa, 2016; Pöchhacker, 2016) but also assist themselves to deduce the sense of the source speech (Seleskovitch, 1978). Albeit the significance of notetaking, based on the corresponding finding in the previous chapter, whether GA can be accompanied by the action of taking notes seemed to be influenced by whether concurrently executing active listening and notetaking would be attainable. It is worth reiterating that cognitive load is additive (Xie and Salvendy, 2000; Paas *et al.*, 2003, 2005; Galy *et al.*, 2012). In CI, as mentioned in 2.1.6, the need to perform active listening and notetaking imposes a high cognitive cost for interpreters because the limited WM capacity is often stretched (Gile, 2009). Considering that GA would occur when one's cognitive load is high, for GA+NT to be performed, interpreters would need to be exposed to a certain level of cognitive load while still being able to accommodate the active listening and notetaking within the limited WM capacity. See the following example:

Example 3

Sentence 1: *Free vaccinations for life-threate<u>ning and infectious diseases are</u> becoming available.*
Sentence 2: The job is very physically demanding.

As previously reported, GA occurred sparsely under the easy speech. Sentence 1 in the above example, according to the mark-ups in ELAN, was the only sentence that witnessed a group of participants (3 out of 6 GA interpreters) averting their gaze from the notepads while maintaining handwriting. Specifically, the three interpreters conducted 4 times GA during the underlined text was uttered. According to the reasoning above, to induce GA+NT, the sentence would impose a cognitive cost that is enough to stretch interpreters' STM/WM to a certain level but insufficient to interrupt notetaking.

An initial observation of Sentence 1 would reveal that the sentence is composed of a Subject (*Free vaccinations for life-threatening and infectious diseases*) – Linking Verb (*are*) – Predicative structure (*becoming available*), and that the sentence is syntactically simple. However, comprehending this sentence might not be as straightforward due to the long distance between the head noun “vaccination” and the linking verb “are”. As mentioned in 2.6.2.4, the cognitive expense is affected by the distance between the word and its obligatory syntactic components, with a longer distance resulting in a higher processing load (Gibson, 2000). In the subject of Sentence 1, the head noun and its obligatory syntactic components are separated by a description that defines the head noun. Aspiring to attain maximum faithfulness, the interpreters would endeavour to thoroughly comprehend the subject before integrating it with the obligatory syntactic elements. However, such an attempt might create an extra burden on interpreters' STM/WM. As mentioned in 2.1.3, scholars acknowledge that STM/WM, the intermediate channel connecting the Sensory Register and LTM, is very limited in capacity and can only hold around 7 information units (Miller, 1956; Baddeley, 2009; Cowan, 2010). The PL, the slave component of WM responsible for holding verbal information in serial order and facilitating articulatory rehearsal, has an even slimmer capacity and can only register about two digits or chunks of input for two seconds (Baddeley and Hitch, 1974; Baddeley, 2009, 2010, 2018). However, a meticulous segmentation (Bovair and Kieras, 1981) of the subject would reveal that processing the subject, depending on the availability of schematic strategies, would propel interpreters to hold a maximum of 7 pieces of information (Free; vaccination; for; life-threatening; and; infectious; diseases), resulting in an increase of cognitive load manifested in GA. Sentence 2, under which no GA was coded, was extracted from the source speech. Like sentence 1, sentence 2 also adopted a Subject (*The job*) – Linking Verb (is) – Predicative structure (*very physically demanding*). However, Sentence 2 is arguably much more cognitively economical to process as the subject (*The job*) seamlessly connects with its obligatory component (*is*). In other words, participants were not affected by the distance between the two syntactic components. Therefore, the cognitive expense for Sentence 2 was much less significant than what was required for Sentence 1 and did not induce GA.

Another possible explanation could be that the compound adjective (life-threatening) recruited greater cognitive resources. Studies have found that compared with monomorphemic words, compounds could induce longer reaction times (Yagoubi, Chiarelli, Mondini, Perrone, Danieli, and Semenza, 2008; Fiorentino, Naito-Billen, Bost, and Fund-Reznicek, 2014). According to Fiorentino *et al*., novel compounds are processed significantly slower than their unstructured counterparts (e.g., tombnote vs tomb note). Yagoubi *et al*. (2008) suggested that processing compounds enlist greater cognitive resources as it is necessary to access the whole word and its components. From the perspective of WFE, the longer reaction time from processing compounds could be stemmed from the possibility that the compounds would function as rare words and, therefore, trigger a longer memory search process. For the participant who performed GA towards the end of the compound adjective, it could be the case that the word was novel and imposed a higher processing cost.

As mentioned above, taking notes while averting gaze from notepads would require available cognitive resources, and interpreters would strive to stay engaged in the action due to the profound benefits of notetaking to the quality of interpreting. Therefore, attempting to resort to GA+NT significantly more often than GA-NT would be reasonable. However, since interpreters' cognitive load was high enough to drive GA, it would seem justified to ask why there would be available resources to underpin notetaking, a task that also can be cognitively expensive. Explanations can be drawn from at least two aspects. To begin with, as exhibited in Example 3, the easy speech was composed of syntactically simple sentences containing common words. The interpreters were competent English users with sufficient linguistic and schematic knowledge to process the speech. Therefore, it would be warranted to argue that the overall cognitive load imposed by the easy speech was under the level where notetaking must be sacrificed to prioritise active listening.

In a nutshell, given the significant bearing of notetaking, interpreters would endeavour to stay engaged with the behaviour. However, taking notes while prioritising active listening depends on sufficient cognitive resources. During the easy speech, interpreters' overall cognitive load was kept at a relatively lower level where GA regulated rare load upsurge. As a result, GA was significantly more frequently accompanied by the action of taking notes.

4.3.3.2 The decrease of GA+NT in Difficult Speech

As previously mentioned, GA level became significantly higher in response to the increased difficulty level of the speech difficulty. Within the scope of investigating the difference between the levels of GA+NT and GA-NT, the descriptive data showed a reversed preference for one type of GA over another. Specifically, GA-NT became more prevalent than GA+NT during the difficult speech. However, inferential analyses did not render the differences statistically significant in either counts or length measurements. So far, regarding interpreters' preferences toward a certain type of GA, this thesis has pointed out that both types of GA have cognitive groundings: GA+NT indicates ample cognitive resources to maintain both GA and notetaking, whereas GA-NT suggests the cognitive load has increased to the level where GA and disengagement from notetaking are required to prioritise active listening. To gain a complete picture of how the fluctuation of cognitive load might affect whether note taking would be maintained during GA, it would be necessary to explain the decrease of popularity with GA+NT.

From the cognitive load perspective, the decrease of GA+NT, or the increase of GA-NT, could indicate a rise of cognitive resources enlisted during active listening and notetaking. The evidence for this notion would primarily stem from the sharp comparison between the difficulty levels of the easy and challenging speech. As shown in 4.1.2.1.4, the easy speech did not contain LFWs and was composed of syntactically simple sentences only. By comparison, the difficult speech exposed interpreters to three types of more challenging sentences: syntactically simple sentences containing rare words, syntactically complex sentences containing common words, and syntactically complex sentences containing LFWs. The sharp increase in speech difficulty would mean that interpreters must devote more resources to comprehending the speech. If maintaining notetaking and active listing would result in a deficit in cognitive resources, notetaking would need to be sacrificed so that comprehending the source speech can continue, a notion that echoes with Gile's Effort Model (Gile, 2009).

4.3.4 Absence of Priming Effect in Hit Rate and GA

The second research question pursued in this experimental study asked if exposing interpreters to a list of LFWs Would create a priming effect where longer exposure would contribute to higher hit-rates and lower GA levels. It was hypothesised that longer exposure would result in better accuracy in rendering the words and witness a lower GA level compared with shorter exposure because the WFE induced by the rare word would be much weakened by repetition priming. However, data analysis did not reveal significant effects in either aspect. The following sections explain the unexpected absence of a priming effect.

4.3.4.1 Absence of Priming Effect in Hit Rate

The absence of priming effect in terms of hit rates can be explained from two aspects, and the first concerns the possibility that regardless of the length of exposure, the LFWs contained in either chunk of the difficult speech remained largely unknown to interpreters. According to Monaghan *et al*. (2017), for rare words to cast an effect, they would need to be encountered a few times to leave a trace in one's mental lexicon. Given the frequency of the LFWs and the fact that the interpreters use English as L2, it would be warranted to argue that some words such as "penchant" (Zipf = 2.87) and "disparage" (Zipf = 1.93) have been rarely encountered by the interpreters. In other words, it is likely that at the time of the priming sessions, some of the LFWs were too fresh to leave a trace in interpreters' mental lexicon. During the priming sessions, although it would be expected that the interpreters endeavoured to rehearse the words in their STM, the slim time window might have deprived them of gaining substantial familiarities. In other words, the 8 minutes prime session barely allowed the interpreters to gain an elementary level of familiarity with the words, l*et al*one the shorter prime session. The evidence to back up the above reasoning would be the observation that the rendition of LFWs under both longer and shorter exposure conditions were often similarly unsatisfactory. See the following example.

Example 4

Source Sentence: I am always a **fervent** person in believing in goodness.

Rendition Version 1 (Longer Exposure): *我一直相信人性本善。*(*Back translation: I always believe that human beings are kind at heart.*)

Rendition Version 2 (Shorter Exposure): *我一直是一個比較相信善良的人。*(*Back translation: I have always been a person who is rather willing to believe in kindness.*)

The above example compares two renditions produced under the longer and shorter exposure conations for the same source sentence. The LFW, "*fervent*" (Zipf = 2.69), is bold. A contrast between the notes produced by the two interpreters revealed that, regardless of exposure time, both interpreters recorded the "*fervent*" as a word that carries a sense of enhanced degree – the interpreter who was primed longer noted down the word as an exclamation mark, whereas the other interpreter quickly wrote down “熱" (could stand for "*passion*” or “*hot*”) moments after the word was uttered. As discussed in 2.6.2, scholars have reached the consensus that the disadvantage of rare words would gradually fade as the familiarity with the words increases (Forster, 1981; Brysbaert *et al*., 2017; Monaghan *et al*., 2017). In the context of the source sentence, if the longer exposure did result in a priming effect, a pathway would have been cleared for the interpreter who was primed longer to attain a quicker and more successful lexical activation process than the interpreter who underwent the shorter exposure condition, which will be reflected in the target speech. However, neither interpreter successfully rendered the LFW. Comparing the back translation against the source sentence would show that the interpreter primed longer for the word "*fervent*" omitted the word in the rendition. Instead, the participant used “*always*”, a word that does not carry a similar meaning with “*fervent*”. By contrast, the interpreter primed shorter for the same word seemed to have registered it as an adjective that defines "*person*", although the original meaning was distorted in the target speech. For both interpreters, the contrast between the ability to note down the word and the inability to interpret it with faithfulness would suggest that the exposure hardly resulted in any priming effect - an argument would echo with the discussion in 2.6.2.4. To reiterate, during active listening, although detecting a rare word would be swift because LFWs typically pose a stand-out effect as new words (Guttentag and Carroll, 1997; Malmberg *et al*., 2002; Yonelinas, 2002; Neville *et al*., 2019), recalling the meaning of a rare word could be cognitively arduous. In the context of the source text, it would be warranted to argue that the lexical retrieval process for the LFW in question was similar at large, suggesting that the priming effect, if any, was insignificant.

Another perspective of explaining the absence of a priming effect regarding hit rate stems from the possibility that the challenging global cognitive load imposed by the task prevented the priming effect from manifesting. Scholars on the subject of cognitive load have accentuated that intrinsic and extraneous cognitive load are additive (Xie and Salvendy, 2000; Paas *et al*., 2003, 2005; Galy *et al*., 2012), implying that a high level of global cognitive load could emerge when either or both types of loads become challenging. In interpreting studies, challenging global cognitive load is perhaps more commonly known as WM saturation. According to Gile (2017), proximity toward WM saturation would temporarily reduce interpreters’ capability. The difficult speech, as discussed in 4.3.2, stretched interpreters' WM capacities by propelling interpreters to engage in arduous active listening and notetaking, contributing to proximity to WM saturation. Based on Gile's suggestion mentioned above, it would be warranted to argue that, when working under the cognitively gruelling situation, the available cognitive resources were too meagre to facilitate lexical retrieval, diminishing the opportunity for a priming effect to be seen. See the following example.

Example 5

Source Sentence: *I wonder if it will be a new normal where parents <u>whom we think should safeguard the children</u> will need to expose them to the Internet to teach them the danger of being* ***credulous***.

Rendition Version 1 (Longer Exposure): *同時，我也很擔心這會成為一個新的常態，那就是父母為了保護孩子們不受騙而阻止他們上網，以此來避免他們受到信任危機。*(*Back translation: In the meantime, I also worry that this will become a new normal, that is, parents, in order to protect their children from tricks, will have to stop them from getting on the Internet so that the children can be free from trust crisis.)*

Rendition Version 2 (Shorter Exposure): 所以有可能在今後會形成一個新常態，那就是是不是家長會故意讓孩子們去接觸這些互聯網上的壞東西，來讓他們親身感受到接觸這些東西的危險。(*Back translation: so perhaps from now on, there will be a new normal, that is, parents might purposely let their children experience the bad stuff on the Internet so that the children could gain a personal experience about the danger of accessing these bad things.*)

The above example compares two target speech sentences performed under longer and shorter exposure conditions. The source sentence was composed in a syntactically complex manner and contained an LFW "credulous" (Zipf = 1.84). It is arguably the case that the example has adequately demonstrated interpreters' loss of sense when the task demand is dire. Initial observation of the sentence would immediately review that the sentence contains two subordinate clauses with the second nested in the first. As previously discussed in 2.6.2.4, syntactic complexity could be a powerful element that hinders active listening (Chomsky, 1996; Caplan *et al*., 1998; Wingfield and Tun, 2007; Seeber, 2011; Frazier, 2013; Ferreira *et al*., 1996; Gibson, 1998, 2000). However, the sharp rise of cognitive load would arguably result from the underlined object-relative clause. Based on Gibson's (2000) DLT, the underlined object-relative clause would have created a long distance between "*parents*" and “*will need*". For interpreters, dealing with the distance meant they had to hold the clause in their memory stores and keep the structure active while waiting for "*parents*” and “*will need*" to be jointed. Keeping a long syntactic structure active in STM/WM could also increase extraneous cognitive load because the participants would aspire to indicate the semantic relations among sentence components in notes. The challenging task of demand is also reflected in the target speech. From the perspective of measuring cognitive load by task performance (Sweller and Chandler, 1994; Sweller, 2011; Schroeder and Cenkci, 2018), a comparison between the back translation and the original speech would immediately suggest that both interpreters experienced high cognitive load when processing the sentence in the example.

4.3.4.2 Absence of Priming Effect in GA Level

Inferential analyses failed to indicate any priming effect in terms of GA level. Specifically, the difference in GA levels measured under longer and shorter exposure conditions did not reach statistical significance. Although the result is unexpected, the reasons for the absence of priming effect in terms of GA would arguably be apparent and echo with the reasoning in the above section. To begin with, similar to the reason previously stated, considering the low frequency of the LFWs, it is possible that, regardless of the length, the exposure did not suffice to lower the difficulty of recognising the words. Owing to the stand-out effect of the newly encountered words (Guttentag and Carroll, 1997; Malmberg *et al*., 2002; Yonelinas, 2002; Neville *et al*., 2019), the participants might have been startled by the utterance of LFWs, triggering an immediate extensive memory search. As mentioned before, words with lower frequency takes longer to process. The cognitive expense associated with the LFWs would be exceptionally high as the interpreters must deal with the constant in-flow of new information. Considering the possibility that neither exposure condition was sufficient to make a difference, GA levels under the shorter and longer exposure conditions would be inevitably not comparable.

Another reason to explain the absence of a priming effect regarding GA levels is rooted in the possibility that the levels of gross cognitive load were similarly high under both conditions. As argued above, if longer exposure time did result in a reduction of WFE for the corresponding LFWs, the combined intrinsic cognitive load from active listening and the extraneous cognitive load imposed by notetaking would have placed the benefits in a futile position. Similarly, although it could be argued that the longer exposure contributed to an easier lexical activation, the liberated cognitive resources from priming were too insignificant to substantially replenish the resource. In other words, exposure to LFWs did not suffice to regulate cognitive overload. In Example 1 in the (4.3.2.1), although each half of the participants were subject to either longer or shorter exposure to LFWs, the variances in the degree of exposure did not seem to have made a difference regarding GA level as a cohort of GA was achieved by 8 out of 16 of the interpreters, among which 6 were primed longer.

4.3.5 GA and Interpreters' Performance

The third RQ pursued in this experimental study concerns testing a critical aspect of the Cognitive Load Hypothesis by exploring if GA would enhance interpreters' performance measured by completeness and accuracy. In this study, the hypothesis for RQ3 was that for easy speech, the performance-enhancing benefit of GA would be limited because interpreters' cognitive load would be kept at a relatively lower level due to the low difficulty of the source speech. For the difficult speech, it was hypothesised that interpreters' performance would positively correlate with GA levels. However, as indicated by the results of inferential analyses, these hypotheses were rejected as there was no significant effect between GA and non-GA interpreters. Surprisingly, however, inferential analysis results suggested that within the group of GA interpreters, sentences where GA happened were rendered significantly worse than sentences that did not within this GA, suggesting that instead of enhancing performance, GA would be associated with performance deterioration in CI. This finding contradicts previous studies that reported that GA would improve concentration and task performance (Phelps *et al.*, 2006; Phelps *et al.*, 2010). The sections below are dedicated to explaining the results.

4.3.5.1 Non-significant Performance Variances under Both Speech

The primary reason why the performance variances between GA interpreters and non-GA interpreters did not reach statistical significance under the easy speech condition could be that the task difficulty for either type of interpreter was similar. Under the easy speech condition, there would be little doubt that the participants were exposed to relatively lower levels of cognitive load. As previously discussed, the participants are competent English users who would arguably have acquired sufficient schematic knowledge to deal with the speech where sentences were short and free from rare words. The easy-to-handle source speech would also contribute to a straightforward notetaking process as arranging information and deducing logic would be straightforward. In other words, from the perspective of the construct of cognitive load, the intrinsic and extraneous cognitive load induced from processing and taking notes from the source speech would be on the lower side, generating ample germane cognitive load with which interpreters could retrieve schematic knowledge from LTM stores in a smooth and timely manner. The low cognitive load under the easy speech condition, apart from resulting in scant GA, would also have limited the variances regarding performance among GA and non-GA interpreters. See the following examples:

Example 6

Sentence 1: *Regular sleeping hours to me is a luxury.*

Rendition from a GA interpreter: *到點上床，睡上完整的好覺，對我來說是相當奢侈的。*

(Back Translation: Going to bed at a fixed time and having a full and good sleep, to me, is rather luxury.)

Rendition from a non-GA interpreter: *對我來說，一個正常時間的睡眠是一種奢侈。(Back Translation: To me, a normal hour for sleeping is a luxury.)*

Sentence 2: *But 18% of local population managed to escape poverty.*

Rendition from a GA interpreter: *百分之十八的當地人已經脫離了貧困。 (Back translation:18% of local population has escaped poverty.)*

Rendition from a non-GA interpreter: *但是百分之十八的柬埔寨人已經脫離了貧困。(Back translation: But 18% of Cambodians have already escaped poverty.)*

Each of the above examples compares the rendition of the same sentence by a GA and non-GA interpreters. It would be warranted to argue that the intrinsic cognitive load that stemmed from processing the two sentences in the example was low due to the simple syntactic structure and absence of LFWs. Comparing the two versions of the rendition of Sentence 1 would reveal that both interpreters rendered the source text with faithfulness. The noticeable variance between the two versions is the added information ("*睡上完整的好覺*" (*having a full and good sleep*)) in the GA interpreter's rendition. However, the addition is not penalised as the marking was based on completeness and accuracy (see 4.1.3.3). For Sentence 2, the non-GA interpreter, who rendered the source speech completely and accurately, outperformed the GA interpreter, who omitted the coordinating conjunction "*But*", which resulted in a mark deduction. Regardless of the difference, the two renditions share remarkable similarities in delivering the sense of the source text.

As evidenced by the sharp increase of GA, the difficult speech that included complex syntax and LFWs, induced much higher levels of cognitive load than the easy speech. According to the cognitive load hypothesis, the increase in GA induced by the high processing load would enhance interpreters' performance (Glenberg *et al*., 1998). However, the inferential analysis found that the performance differences between GA and non-GA interpreters were not significant. Similar to the reason discussed above, it could be suggested that the absence of significant effect was due to the possibility that the difficult speech greatly challenged both types of interpreters.

Example 7

Sentence 1: *The boy* ***whom we think highly of does not hold any rancour towards*** *the misconduct of his mate*, ***which would have been a better situation because instead***, *he has relinquished his pursuit of friendship.*

Rendition from a GA interpreter: *並且是由於他的朋友做出了一些不當的行為，所以他最後就放棄了這樣的一段友誼。 (Back Translation: Also, because the misconduct that his mate has done, he, in the end, relinquished the friendship.)*

Rendition from a non-GA interpreter:*就是他朋友這次不端的行為導致了他產生這些情緒，使得他認為友誼是一件不值得他再去追求的事情。(Back Translation: It is the misconduct of his mate that caused him to have these feelings and made him believe that friendship is something that is no longer worthy of his pursuit.)*

Sentence 2: *While it brings excitement to the tips of our fingers, the Internet has also introduced an environment that I would describe as "if you open your window, there will be flies in your house".*

Rendition from a GA interpreter: *我們通常只要動動手指就可以看到很多有趣的新聞，也可以感覺到環境的變化。(Back Translation: We often only need to move our fingers to see much exciting news and to feel the change in the environment.)*

Rendition from a non-GA interpreter: *你可以從指尖感受到互聯網所帶給我們的興奮，而且它也對我們的環境帶來了改變。(Back Translation: You can feel from the fingertip the excitement that the Internet brings to us. It also brought changes to our environment.)*

Each of the above two examples compares the rendition of a sentence by a GA interpreter and non-GA interpreter when the speech was difficult. Both sentences contain LFWs and are syntactically complex. The underlined source text represents the moment when the GA interpreter performed GA-NT. Compared with the sentences extracted from the easy speech, the above two sentences arguably taxed interpreters' WM to a much higher degree because of the high intrinsic cognitive load resulting from processing LFWs (Forster, 1976, 1981; Scarborough *et al*., 1977; Mutter and Hashtroudi, 1987; Keuleers *et al*., 2010; Brysbaert *et al*., 2018; Vitevitch *et al*., 2018; Neville *et al*., 2019; Popov and Reder, 2020) and complex syntactic structure (Ferreira *et al*., 1996; Gibson, 1998). The challenging source speech would also mean that notetaking consumed more effort than when the source speech was easy, meaning that extraneous cognitive load was also kept at a higher level. Although the high cognitive load was not entirely reflected in GA, as 6 out of 16 interpreters did not resort to the behaviour, it is manifested in the deteriorated performance that usually indicates saturation or overload (Gile, 2009). It should be noted that the renditions of both sentences, regardless of by GA or non-GA interpreters, were far from satisfactory. For Sentence 1, both interpreters omitted a significant amount of information. Specifically, the two relative clauses (in bold) were omitted from both versions, damaging the faithfulness of the target speech. Similarly, for Sentence 2, substantial omissions were observed for both GA and non-GA interpreters.

4.3.5.2 GA and Performance Deterioration

As previously discussed, GA seemed to be associated with performance deterioration instead of facilitating better performances. Specifically, within the group of GA interpreters, regardless of speech difficulty, sentences where GA happened were rendered significantly worse than sentences that did not attract GA. This finding contradicts previous studies where GA was found to improve concentration and comprehension (Phelps *et al*., 2006). In Phelps *et al*.'s study, participants who were 5-year-old children were asked to think of answers for verbal and arithmetic questions. By encouraging participants to look away, which facilitated the re-routing of cognitive resources from looking at a researcher's face to thinking of the answer, participants showed better concentration and comprehension, leading to more satisfactory performances. However, the current study adopted a vastly different experimental design: the material is more challenging than verbal or arithmetic questions, participants are processing L2 while taking notes, instead of providing simple answers, participants in this thesis were commissioned to produce target speech faithfully. It would be warranted to suggest that the task demand in this thesis is much higher than Phelps *et al.*'s study. It would be irrefutable to suggest that compared with answering verbal or arithmetic questions, CI would impose a significantly higher cognitive load. As previously discussed, CI is a fundamentally difficult activity (Nigro, 2015) as it involves active listening and notetaking, both of which could expose interpreters to cognitively dire situations (Cutler and Clifton, 1999; Piolat *et al*., 2005; Jansen *et al*., 2017). Therefore, the association between GA and performance deterioration could be attributed to the possibility that the high cognitive load imposed by CI broke through interpreters' cognitive limits. In other words, GA can indicate cognitive overload. See the following examples.

Example 8

GA Sentence: *The pandemic will push 49 mill<u>ion people</u> into extreme poverty this year.*
Rendition: *大瘟疫會讓百萬計的人辛苦謀生。(Back translation: The pandemic will make life for millions of people miserable.)*
Non-GA Sentence: *The fragile houses are ruined easily by rain, storms, and earthquakes.*
Rendition: *他們的房子經常受到暴風雨、地震的摧殘。(Back translation: Their houses are often battered by storms and earthquakes.)*

The example above compares the rendition of a GA sentence and non-GA sentence by the same interpreter who exhibited GA during the utterance of the easy speech. A single occurrence of GA was coded during the utterance of "*million people*" (underlined). The comparison between the two back translations would show that the interpreter barely rendered the sentence with faithfulness, whereas the non-GA sentence was interpreted with much better faithfulness. Based on the above reasoning, the interpreter's cognitive system was greatly challenged when GA happened in the first sentence, leading to omissions and inaccuracies. As previously discussed, processing numbers could require a considerable volume of cognitive resources (Lederer, 1982; Alessandrini, 1990; Mazza, 2008). Although it could be argued that students who are professionally trained to become interpreters would handle "49 million" easily, the individual differences in processing capabilities and skills should not be overlooked (Daneman and Carpenter, 1980; Unsworth and Engle, 2007).

Example 9

GA Sentence: *The boy who leery is what he is considered has lost his* ***prudence*** *when he brought some fraudulent tutorial videos that* ***allege****d to prepare you to become a* <u>*tycoon*</u> *in the stock market.*

Rendition: *他當時失去了他一貫的小心謹慎，然後看了一個視頻。就是這個視頻，他被騙了。 (Back translation: He lost his usual caution, and then he watched a video. Because of the video, he was tricked.)*

Non-GA Sentence: *What I find* ***besetting*** *is not just the fact that my son was* ***coaxed*** *which led to a financial loss, but that he has already experienced one of the darkest sides of human beings*.

Rendition: 在這背後我發現了一件非常令人不安的事情，那就是不僅僅是因為我兒子被騙了然後損失了錢財，還是因為我兒子他經歷了人性最黑暗的一面。*(Back translation: Behind the incident, I found something that is very besetting, that is, not only because my son was coaxed, which led to a loss of money, but also because my son experienced the darkest side of human nature.)*

The above example compares the rendition of a GA sentence and a non-GA sentence by the same interpreter who resorted to GA during processing the difficult source speech. The underlined text in the GA sentence represents the moment the interpreter looked away from her notepad. The similarities between the two sentences include length, the number of information segments, and sentence difficulty – both are syntactically complex and contain LFWs (in bold). Comparing the back translations against the source speech would make it apparent that the performance for the non-GA sentence significantly surpassed the performance of GA sentence. Whereas the rendition of the GA sentence suffered incompleteness to an alarming degree, the rendition of the non-GA sentence almost faithfully reflected the sense of the source speech owing to its excellent completeness and accuracy, resulting in higher marks than the rendition of GA sentences.

The association between GA and performance deterioration could be caused by cognitive overload. Specifically, in the context of the difficult speech, the performance-boosting benefits of GA could not be observed due to the possibility that interpreters who averted their gaze experienced cognitive load dramatically beyond their STM/WM limits. In example 9, there would arguably be little dispute that the GA interpreter, compared with the non-GA interpreter, due to individual differences in memory capacity (Daneman and Carpenter, 1980) or exposure to L2 (Cop *et al.*, 2015), experienced a more cognitively dire situation evidenced by the substantial omission. According to Gile (2009), interpreters would suffer performance deterioration when WM resources are depleted. Based on DLT proposed by Gibson (2000), the two relative clauses, especially the object-relative clause in Sentence 1, would have imposed a hefty load on interpreters' STM/WM systems due to the necessity of holding them in memory stores before integrating with obligatory syntactic heads. Therefore, processing source sentences, as shown in Example 9 might have created the following situation for GA interpreters: their cognitive systems registered high cognitive load and reacted accordingly by propelling the interpreters to avert their gaze from their notepads to release and channel some resources to comprehending the utterance, but the freed resources were too insignificant compared to the soaring cognitive load, and therefore were insufficient in either decreasing cognitive load or improving comprehension, hence the inferior performances.

4.4 Conclusion

4.4.1 Overview of study

This experimental study was set out to verify the previous findings obtained from the corpus study and further explore the intricacy of GA in CI setting. Specifically, the study investigated the following aspects.

First, the study investigated whether an increase in speech difficulty would result in a higher GA level. Two speech with low and high difficulty, respectively, were used in the experiment. Results suggested that GA was significantly more frequent and lasted significantly longer on average when the source speech was more difficult. This finding confirmed GA's function as a high cognitive load indicator in CI settings.

The study also attempted to explore whether exposing interpreters to the rare words about to be experienced would reduce GA level and improve rendition of the words. Towards this aim, repetitive priming was introduced and manipulated. The results suggested that priming did not affect either GA level or rendition accuracy.

Finally, the study investigated whether the performance-enhancing benefits reported in other studies would sustain in CI setting. A meticulous approach was used to segment the source speech for making. The performance score of GA interpreters and non-GA interpreters and GA sentences and non-GA sentences were compared. The result suggested that regardless of speech difficulty, the performance variances between GA and non-GA interpreters were not statistically significant. However, surprisingly, within the group of GA interpreters, sentences where GA occurred were rendered significantly worse than sentences that did not attract GA. The result would indicate that GA would indicate cognitive overload or performance failure in CI.

4.4.2 Strengths and Limitations

4.4.2.1 Strengths

The study has the privilege of enjoying a few strengths. To begin with, this experimental study made original contributions to knowledge by reporting that GA is associated with performance deterioration and highlighting the possibility that GA, as well as a cognitive load indicator, also indicates cognitive failure in cognitively dire situations such as CI. The strength of the study can also be drawn from a methodological perspective. First, the composition of the source speech considered a wide range of elements. The speech content and the approach with which it was composed would offer valuable experience to future studies. Also, it would be warranted to argue that the data analysis has shown much rigour and robustness. Last, using smartphones to record interpreters' eye movements and doing the experiment remotely shows remarkable resilience and adaptability in response to a difficult situation.

4.4.2.2 Limitations

Like all studies, this experimental study is not free from shortcomings. The most poignant shortcoming would be the inaccessibility to eye trackers, which means that some data, such as pupil dilation, could not be obtained. However, the shortcoming is accepted as the pandemic is beyond anyone's control. It would be warranted for the author to feel lucky that the study was redesigned promptly. Another shortcoming would be that the sample size was small. However, it is reconciled as there were not enough qualified candidates.

Chapter 5. General Discussion

As previously mentioned in Chapter 3 and 4, findings that bear overarching significance will be put together in this chapter to be discussed. These findings concern the comprehensive reason why GA would occur during CI active listening and notetaking, considerable individual differences in all aspects of GA, and whether GA functions as a cognitive load indicator or a cognitive failure indicator in the setting of CI. the following sections are dedicated to discussing these three aspects.

5.1 CI as a GA-Inducing environment

Both chapters 3 and 4 have confirmed that GA would occur in response to high cognitive load, a finding in line with previous studies (Glenberg *et al*., 1998; Doherty-Sneddon *et al*., 2002; Phelps *et al*., 2006). It has also been found in the two studies that GA level correlates with the level of speech difficulty. Although it is irrefutable that some textural features, such as LFWs and complex structures, are mighty in driving up cognitive load, attributing the reason for GA to textual features alone would be an oversimplified perspective because it would be fair to suggest that cognitive load CI is imposed from several sources.

5.1.1 A Stressful Environment

In general, it would be fair to suggest that CI creates an environment where peripheral factors could result in high cognitive load. The core reason for this suggestion stemmed from the idea that stress and anxiety can significantly increase people's cognitive load levels (Paas *et al*., 2003; Cen and Chang, 2009; Galy *et al*., 2012; Chen *et al*., 2016; Bong *et al*., 2016). In the corpus study, the environment where CI was performed was final examinations, during which the interpreter was put under the spot in the room with two examiners who would essentially decide whether the students could continue this/her study as planned. Feeling judged and lacking control would result in enormous stress and anxiety (Chen *et al*., 2016). By comparison, in the experimental study, participants did not interpret for their future studies. Instead, they were helping a senior student to conduct research. Also, the experiment occurred in their bedroom, which they call home. Therefore, it should be fair that participants in the experimental study, compared with their counterparts in the corpus study, experienced much less extraneous cognitive load. From this perspective, there is a possibility that one of the reasons why some interpreters did not perform GA during the experiment is that the extraneous cognitive load was very low, contributing to a more manageable consumption of WM resources.

5.1.2 Physical Efforts

Another reason why CI is GA-inducing is that it can be physically exhausting. Studies have found that tiredness is another source of extraneous cognitive load, and it can make people more sensitive to factors that induce high cognitive load (Zagermann, Pfeil, and Reiterer, 2016; Herbig, Düwel, Helali, Eckhart, Schuck, Choudhury, and Krüger, 2020; Souchet, Philippe, Bourdeaux, and Leroy, 2022). In the corpus study, interpreters performed CI standing up. During the short period of the examination, the interpreters had to maintain a standing-up position in an upright and respectable way because presentation carried some weight in the final score. It would be fair to argue that some physical efforts were consumed in maintaining the position. Another activity that consumed physical effort would be taking notes while standing up because interpreters had to hold the notepad in one hand during writing. Although one could argue that a notebook and a pen would be light to carry, the muscle in the arm and hand must stay engaged for the duration of the task. Admittedly, the physical efforts involved in standing up and taking notes for 10 minutes should not deplete the stamina of a young and healthy person. However, the possibility that physical efforts would make interpreters more susceptible to the factors that increase cognitive load should not be overlooked when trying to understand CI as a GA-inducing situation.

5.1.3 L2 Active Listening

Although it will be oversimplified to consider language processing as the only reason to drive GA, there should be little doubt that language processing rests at the centre of why GA would occur in CI. In previous sections, examples have shown that processing LFWs and syntactically complex structures burden one's cognitive system because the costs of the two factors are simply onerous. It would be warranted to argue that the cognitive expenses spent in L2 active listening can be much higher.

It is well documented in the literature that there is a disadvantage associated with L2 processing. L2 users, regardless of their language proficiency, process the language in a less automatic and more effortful manner (Clahsen and Felser, 2006a; Piolat *et al.*, 2008). As mentioned in 2.6.2.2, CI active listening compels interpreters to resort to both bottom-up and top-down processing to attain faithfulness (Setton and Motta; 2007; Gile, 2009; Pöchhacker 2002, 2015). While bottom-up processing helps construct meaning in a gradual and accumulative fashion (Cutler and Clifton, 1999; Buck, 2001), top-down processing requires utilising schematic and contextual knowledge stored in LTM (Marianne Celce-Murcia, 1995a;

Tsui and Fullilove, 1998; Buck, 2001; Vandergrift, 2011). Studies have found that active listening can be more cognitively challenging for L2 English users. At the word level, L2 users are more vulnerable to phonological interferences (Broersma and Cutler, 2008) and WFE (Cop *et al*., 2015; Brysbaert *et al*., 2017), meaning that recognising rare words can be extra challenging. Sentence parsing for L2 users is also more cognitively demanding because, compared with native speakers, L2 users may not receive substantial support from schemata due to less language exposure (Clahsen and Felser, 2006b; Jiang, 2004, 2007; Hopp, 2015). L2 users, therefore, more often engage in shallow processing constrained by surface cues such as lexical-semantic information. When the syntactic complexity of a sentence becomes challenging, L2 users, or interpreters who work from English to their mother tongue, would find it exceptionally challenging. In short, active listening would impose a higher intrinsic cognitive load on interpreters working from L2 to L1.

5.1.4 Summary

In summary, CI forms an environment where several factors contribute to the sharp increase of cognitive load. From the perspective of task demand, CI interpreters must exhibit unmitigated concentration on processing the source speech and recording information in their notepads. Active listening can entail an arduous process where interpreters rely on substantial support from linguistic and general knowledge to deal with challenging factors such as complex syntax and rare words, imposing a high intrinsic cognitive load on interpreters. The cost associated with language processing is arguably higher for interpreters who work from L2. From an ergonomic point of view, CI can also expose interpreters to anxiety and stress, two factors that put extra burdens on interpreters' cognitive systems. It would be fair to suggest that CI expose interpreters to high cognitive load, making it a GA-inducing environment.

5.2 Individual Differences

Another striking feature observed in the two studies is the considerable individual differences across all aspects measured in the two studies. As mentioned in the previous section, while some interpreters resorted to GA occasionally, some performed GA with much higher frequency. In the experimental study, several interpreters did not even show GA at all. The following section is dedicated to discussing the drastic individual differences from the perspective of memory skills, personal habits, and language proficiency.

5.2.1 Individual Differences in WM Capacity

Many studies have suggested that individual differences in memory skills can be responsible for variances in task performance (e.g., Daneman and Carpenter, 1980; King and Just, 1991; Kiewra and Benton, 1988). Compared with people with smaller WM capacities, individuals with higher WM spans are better enabled in various tasks. For example, high-span readers are found to comprehend sentences with higher accuracy (Daneman and Carpenter, 1980). King and Just (1991) reported that WM capacity plays a crucial role in processing syntax – individuals with higher capacity process syntax more efficiently. In terms of notetaking, it was reported that people with better WM skills could include more words, complex prepositions, and main ideas (Kiewra and Benton, 1988). It would be fair to suggest that, in these cognitive tasks, people with better WM skills accommodated the task in a relatively easy manner, meaning that they did not suffer from high cognitive load. By contrast, individuals who were less enabled by the WM system did not have enough resources to handle the task demand and therefore were outperformed. Since task performance variances were used to indicate the discrepancies regarding WM capacity, it would seem warranted to wonder if the same rationale can be applied to GA. Both studies in this thesis demonstrated dramatic individual differences regarding GA levels. Since GA is considered a cognitive load indicator, it would be fair to argue that the participants who averted their gaze frequently experienced more frequent episodes of high cognitive load.

Perhaps the contrast would be more suitably made in the context of the experimental study where all participants were exposed to the same conditions but showed dramatically different GA levels – some participants did not conduct GA, whereas some resorted to the behaviour with high frequency even when the speech was easy. Following the avenue that better WM skills facilitate better performance in reading, syntax processing, and notetaking, it would be fair to suggest that participants who resorted to GA more frequently are equipped with smaller WM capacities. By contrast, participants who did not avert their gaze even when the source speech became challenging, therefore, would be likely to enjoy a more accommodating WM system. Although it will not be possible to confirm the suggestion within the scope of this thesis, as participants' WM spans were not tested, it is well grounded in the previously reported correlation between capacity and performance. The suggestion would also shed light on those moments in ELAN when GA seemed to have occurred to some participants for no apparent reason or when GA did not happen at all.

5.2.2 Individual Differences in Notetaking Habits

There is also a chance that the vastly different GA levels exhibited by participants in the two studies were due to varying notetaking habits. Some scholars and professional interpreters often refer to CI notetaking as a habit (Gillies, 2019; Zhang and Wu. 2017), implying that individual features are involved in the activity. As mentioned in 2.6.3.2, Albl-Mikasa (2017) highlighted individuality as one of the three basic principles of CI notetaking, accentuating that CI notes should represent interpreters' personal language. It would be warranted to argue that individuality also manifests in interpreters' commitments to the activity. Given the gravity of CI notetaking, it would be expected that interpreters ~~would~~ endeavour to take as many notes as possible. Indeed, for the encoding and storage benefits of notetaking to be salient, it would require the interpreter to condense a substantial amount of key information (Gary and DiVesta, 1972; Kiewra *et al*., 1991; Kiewra, 2016; Poverly and Wolf, 2019). However, interpreters, especially trainee interpreters, are often found to exhibit over-reliance on notetaking (Chen, 2022), even when the efforts invested in notetaking should be reduced to prioritise comprehension (Gile, 2009; Gillies, 2019). As mentioned in 3.3.2.1, notetaking naturally requires eye-hand coordination (Ujbányi *et al*., 2020), and GA during notetaking would expose interpreters to an unnatural situation in which some interpreters might not be willing to find themselves. Therefore, it is also possible that participants who exhibited less or no GA were too committed to the notetaking.

5.2.3 Individual Differences in Language Proficiency

The last aspect regarding individual differences would arguably concern language proficiency. As mentioned in 2.6.1, CI is a complex linguistic task that involves language comprehension and translation (Rusell, 2005). It has also been discussed that successful active listening particularly relies on interpreters' skilled language proficiency ranging from vocabulary size to parsing complex syntax. Many studies have demonstrated that the degree of language exposure positively correlates with language proficiency (Kalia, Wilbourn, and Ghio, 2014; Cop *et al*.,2015; Brysbaert *et al*., 2017; Field, 2019; Wilde, Brysbaet, and Eyckmans, 2020). Kuperman and Van Dyke (2013) suggested that for L2 users, inadequate language exposure often results in a smaller vocabulary size, indicating that they would be more prone to WFE. In terms of speech comprehension, it has been found that L2 listeners with less-skilled language proficiency may not be apt at bottom-up processing and would heavily rely on semantic instead of syntactic knowledge to comprehend the speech (Celce-Murcia, 1995a; Tsui and Fullilove, 1998), which is often inadequate for fulfilling active listening (Cutler and Clifton, 1999). Therefore, the discrepancy in GA levels among the participants from the two studies could also be linked to considerable variances in language proficiency. In specific, some interpreters conducted GA frequently because their linguistic knowledge often fell short and resulted in a high processing load, whereas others who showed less GA benefited from more skilled language proficiency and resolved challenges with ease.

5.2.4 Summary

Considerable individual differences have been observed as an overall feature of data across all GA measures. The fundamental reason that underpins such difference lies in participants' WM skills. Interpreters with more enabling WM systems would simultaneously manoeuvre through active listening and notetaking tasks. By comparison, interpreters with smaller WM spans often find themselves needing more cognitive resources, which propelled them to conduct GA. It is also possible that some interpreters chose not to look away from notepads so that they could stay committed to recording more notes. Finally, it is also possible that interpreters with less skilled language proficiency would need more GA to reroute some resources to comprehend the source speech in a top-down manner, whereas interpreters with higher language proficiency would have acquired more automatic processing mechanisms and therefore relied less on GA.

5.3 GA As a Cognitive Crisis Indicator

5.3.1 GA Indicates High Cognitive Load

The thesis has found that GA levels positively correlate with cognitive load, as higher speech difficulty often led to more frequent and more prolonged GA. The finding constitutes one of the earliest proofs that introduced GA to the field of Interpreting Studies as a reliable cognitive load indicator. The cognitive load indicating role of GA in CI settings would arguably be perfectly explained based on the readily proven rationale. According to the pioneering scholars (Glenberg *et al.*, 1998; Doherty-Sneddon *et al.*, 2002; Doherty-Sneddon and Phelps, 2005; Phelps *et al.*, 2006), the behaviour would happen at critical moments when a high level of cognitive load is experienced. By averting the gaze, cognitive resources invested in distractive stimulus can be rerouted to tasks more at hand. During CI active listening and notetaking, the main task is to comprehend the source speech to the maximum level so that the target speech can faithfully communicate the sense of the speech and the speaker. However, as previously discussed, CI active listening is a cognitively arduous task requiring unmitigated attention and sophisticated linguistic knowledge. Given the WM system's limited capacity, interpreters often face formidable cognitive load. During CI, interpreters typically take notes to alleviate memory pressure (Pöchhacker, 2016). Interpreters' notes contain rich information that often bears significance to the faithfulness of the target speech. However, the stress-alleviating task enlists cognitive prices as it is executed under dire time pressure (Piolat *et al.,* 2017) and involves planning and decision-making (Jansen *et al.*, 2017). Therefore, CI is an environment where interpreters constantly face high cognitive load. As previously discussed, speech processing is susceptible to challenges such as LFWs and complex syntax, which are known to increase processing load. When interpreters are exposed to high cognitive load, they would require extra resources to maintain comprehension and notetaking. However, since cognitive resources are in limited supply, it would seem natural for interpreters to reroute resources from monitoring the information-rich notebook to active listening.

5.3.2 GA Indicates Cognitive Crisis

This thesis took a new perspective to look at GA, that is, to suggest GA as a cognitive crisis indicator. The above pioneering scholars on GA suggested that GA can enhance performance (e.g., Glenberg *et al*., 1998; Phelps *et al*., 2006). The rationale is clear: the resources channelled from distractions to the task at hand would improve the processing capability, leading to better concentration. However, the performance-enhancing benefits of GA were proposed in studies where 5-year-old children answered verbal and arithmetic questions. Until this study, GA was not tested in a CI setting. The experimental study found that the performances of GA interpreters were no better than those of non-GA interpreters. However, GA sentences were rendered significantly worse in the group of GA interpreters than non-GA sentences, regardless of speech difficulty. Therefore, instead of boosting performance, GA was associated with performance deterioration.

To explain the inconsistency, it would be vital to review the different settings where GA was observed. It would be fair to argue that CI settings, as described above, are more cognitively dire than situations where questions are answered. Therefore, it would be warranted to argue that, in the context of this thesis, interpreters' cognitive systems were preloaded before speech difficulty was experienced. Therefore, when interpreters conducted GA, the cognitive load could have already broken through their limits, and the freed cognitive resources from monitoring notetaking might be too insignificant to rein in the soaring processing load. As a result, sentences that attracted GA happened to be the moments when interpreters' cognitive systems were overloaded, which would explain why GA sentences were rendered significantly poorly. In other words, interpreters' GA could be seen as a "cry for help". Based on the data and the reasons above, it would be fair to nominate GA as an indicator of cognitive crisis.

5.3.3 Summary

GA often indicates high levels of cognitive load and occurs at critical moments when cognitive resources are required. This thesis, for the first time, confirmed the indicative role of GA in interpreting studies. CI Create an environment where interpreters are exposed to high cognitive load imposed either from processing the speech, notetaking, or stress and anxiety. In response to high cognitive load and to prioritise comprehension, interpreters often disengage from the interfering notepads to reroute cognitive resources from monitoring handwriting to comprehending the speech. This load-regulating function of GA has been found to bear a performance-enhancing benefit as the freed resources will top up the capability of focusing on the task at hand. However, the benefit was not found in this thesis. Instead, GA was found to be associated with performance deterioration. The possible reason is that when interpreters conducted GA, they experienced cognitive overload, which could not be alleviated by the resources freed by GA. Therefore, in CI settings, GA could indicate cognitive crisis.

5.4 Issues Concerning Operationalising Speech Difficulty

One of the most prominent findings from this thesis is that speech difficulty significantly affects GA levels. Specifically, increased speech difficulty would result in higher GA levels. It would be warranted to suggest that sentences constructed with either complex structures (e.g., embedded clauses) or rare words (e.g., pseudo) or both are more challenging than sentences containing simple structure and familiar words because higher syntactic complexity (e.g., Just and Carpenter, 1992; Miyake *et al*., 1994; Wingfield *et al*., 2003) and lower word frequency (e.g. Keuleers *et al*., 2010; Brysbaert *et al.*, 2018; Neville *et al*. 2019) consumes more cognitive resources. However, operationalising speech difficulty in the study faces a question: Would some syntactically complex sentences from the difficult speech actually be challenging for interpreters?

For instance, in 3.1.5.3, a sentence “*Does it make any difference to your class calculator when you hear ‘how do you do’ versus “pleased to meet you’?* is deemed as a complex sentence, but it might be argued that “*how do you do*” and “*pleased to meet you*” could be processed with ease as they are idiomatic expressions that can be processed as chunks. Although the sentence would be effortlessly understood in conversational settings, it can be challenging in

CI settings. A standard CI setting requires interpreters to render the speech by chunks instead of sentences, compelling interpreters to take notes, an activity where the flow of new information is constant. In this sense, if interpreters could render a speech sentence by sentence, then notetaking would not be required, and processing the sentence in the example would indeed be much more manageable. Also, in the thesis, all participants worked from L2 to L1. As discussed in 2.6.2.1, L2 processing consumes more cognitive resources than L1 processing. Therefore, the speech difficulty is operationalised from the perspective of interpreters - although some complex sentences in the speech would not be as challenging in a different situation or for native speakers, they would tax interpreters’ cognitive systems in CI settings.

Chapter 6 Conclusion

6.1 Introduction

This PhD project investigated GA in the context of CI. The project made original contributions to knowledge by (1) introducing GA as a reliable cognitive load indicator to interpreting studies; (2) nominating GA as an indicator for cognitive crisis.

The intricacy of GA was robustly investigated via two studies: an initial corpus study and an experimental study. As this thesis constitutes the first study to explore GA in CI, it would be paramount to establish the fact that interpreters would perform GA during CI active listening and notetaking. Towards this goal, a corpus was built up and coded the Coopers study also explored possible reasons why GA is triggered. an experimental study was subsequently conducted to verify the findings obtained from the corpus study in a controlled environment and to further and to investigate further the intricacy of GA Putting the two studies together, this thesis is geared towards the following aims:

> *Aim 1: To explore GA in CI settings*
> *Aim 2: To explore the reasons that trigger GA*
> *Aim 3: To explore if repetition priming would result in lower GA levels and better accuracies.*
> *Aim 4: To test if GA bears performance-enhancing benefits in CI*

Aim 1 and 2 were first pursued in the first study where variables were not controlled due to the nature of corpus studies. The findings yielded from pursuing the two the aims were tested in the experimental study where GA was explored in a controlled environment. Aim 3 was explored in the experimental study. A number of RQs were developed in pursuit of the above aims. Aim 1 pertains to RQs that concern whether GA would be observed among interpreters (RQ1) and whether notetaking would be forfeited during GA (RQ2, RQ4). Aim 2 comprises questions that set out to explore possible GA triggers from the source speech (RQ3, RQ4). Aim 3 involved the question if exposing interpreters to LFWs contained in the source speech beforehand would lead to lower GA level and better accuracy (RQ5) Aim 4 was pursued by answering the question whether GA contributes to better performance (RQ6).

In the first study, a corpus was built up and coded so that GA as a behaviour can quantitively described and compared. The raw material used to construct the corpus was 30 video recordings of CI final examination (English to Mandarin) for Stage 1 trainee interpreters from the MA Translating and Interpreting Programme at Newcastle University from 2013 to 2016. Construction of the corpus was achieved via a series of software, among which ELAN was the platform where videos were coded. Mark-ups were designed to indicate GA and GA type (whether GA is accompanied by notetaking). Other data type, such as textual features were also coded. The data was initially quantitively analysed. Inferential analysis was conducted to test findings from quantitative analysis.

The second study was conducted to verify the findings obtained from the corpus study in a controlled environment. The verification bears significance to this thesis as the corpus where the findings were obtained contained small sample size and various source speech. The experimental study also aimed at exploring whether GA would bear a performance-enhancing benefit as suggested in other studies. 16 participants were recruited to perform a CI task. A source speech where variables were controlled or manipulated was composed. To abide by social distancing rules, the experiment was conducted online. Interpreters' eye movements were recorded with a smartphone. Interpreters' Livenotes were obtained via Livescribe Smartpen. The video recordings of the 16 participants were coded in ELAN. Data were analysed quantitively and inferentially to draw reliable conclusions.

6.2 GA in CI (Aim 1 and 2)

The corpus study found that interpreters would resort to GA during CI active listening and notetaking, but the behaviour was subject to considerable individual differences. The finding was verified in the experimental study which reported more salient individual difference as a number of participants did not conduct GA. In the corpus study, an array of textual features was found to be related to GA, but the analysis suggested that GA frequency and average GA length were positively correlated with high speech difficulty defined by complex sentence either structure or LFW or both. The experimental confirmed that high speech difficulty had a significant effect on GA frequency, and GA levels were significantly higher when the interpreters processed the difficult speech which contained variables that increase difficulty than the easy speech where only simple sentences and common words were uttered. Average GA length trigger from processing complex sentence containing LFWs was found to be significant longer than processing simple sentences with LFWs or complex sentences without

LFWs. The correlation between GA levels and speech difficulty confirmed that the reliability of GA as a cognitive load indicator sustains in the field of interpreting studies

GA was found to be subject to considerable individual differences in terms of timing, frequency, average GA length, and if the behaviour is accompanied by the action of taking notes. The vast individual differences among the participants are primarily stemmed from the variances in participants' WM skills. Interpreters with bigger WM spans are believed to rely less on GA because high-span participants would be able to accommodate the task relatively easily whereas interpreters with smaller spans would often be in deficit in cognitive resources, which would manifest as GA. It also possible that the individual differences were down to over reliance on notetaking. Additionally, differences in language proficiency might also have contributed to the dramatic
variances.

6.3 Absence of Priming Effect (Aim 3)

Despite the longer and shorter priming session in the experimental study, there was no priming effect found in GA levels or hit rates. It could mean that exposing interpreters to LFWs that about to be uttered in the speech did not reduce WFE because the exposure time (4 minutes for shorter exposure and 8 minutes for longer exposure) was too insignificant for participants to gain familiarities with the words. Another perspective would be that the interpreters did gain some familiarities with the rare words, but the overall cognitive load was too profound for the priming effect to manifest.

6.4 Absence of GA’s Performance-Enhancing Benefit (Aim 4)

Previous studies have suggested that GA facilitates better concentration and comprehension, and considering the two actions are often needed in cognitive tasks, GA has been hypothesised to boost performance. In the experimental study, rigorous efforts were invested in meticulously segmenting the source speech so that interpreters’ performance can be assessed more objectively. If the performance-enhancing benefit of GA were to sustain in this study, GA interpreters would have significantly outperformed non-GA interpreters. However, the results suggested that GA interpreters performed no better than non-GA interpreters. Also, surprisingly, within the group of GA interpreters, sentence that attracted GA were significantly worse rendered than sentences that did not witness GA.

6.5 Revisiting GA

Cognitive psychologists have shown sustained interest in human eyes. As mentioned in 2.3.1, human beings start to engage in gazing behaviours early in life (Farroni *et al.*, 2002). In cognitive activities, gazing patterns can indicate the fluctuation of cognitive load, such as fixation (Just and Carpenter, 1980), Saccades (Duchowski, 2017), and GA (Doherty-Sneddon *et al.*, 2002). Pioneering scholars on GA have established that GA is an overt strategy to regulate high cognitive load, and therefore GA is seen as a cognitive load indicator. So far, the exploration of GA has mainly focused on conversational settings where subjects are instructed to answer questions of varying difficulty (Phelps *et al.*, 2006). The application of GA was extended to video-mediated settings (Doherty-Sneddon and Phelps, 2005), where participants answered questions from the interlocutor on the monitor. The authors found that subjects would also avert their gaze from the face on the monitor when the processing load was high.

The rationale of GA is deeply rooted in human cognition. Decades ago, Just and Carpenter (1980) suggested eye-mind coordination - people process what they see. By "process", it implies that there is a cognitive cost. Therefore, if the information captured by the eye is complicated, the cognitive price associated with the processing would also be high. The eye-mind processing touches the core reason for GA - the eye are formidable tools for picking up information, and the human face is full of it. The literature has suggested that the human face is attention-grasping because of the rich signals that can be sent from it (Bruce and Young, 2012) and that people would be more inclined to look at faces than other objects. Since faces are informative and attention-grabbing, face-to-face conversations can carry a cognitive price because people would always try to decode faces (Russell and Lavie, 2011). When high cognitive load occur, maintaining the gaze at a face would be difficult due to the limited supply of cognitive resources. To prioritise the task at hand, people often look away from faces.

More than two decades have passed since Glenberg *et al.* (1998) processed the Cognitive Load Hypothesis. With this project, GA has been brought to a new territory. It may seem peculiar to bridge GA and CI. After all, CI does not usually involve face-to-face conversations. However, the similarities between conversations and CI would become apparent when CI notetaking is brought to the party. There would be little doubt that CI notetaking bears irreplaceable significance to the quality of interpreting. In a usual CI setting, an interpreter must process the speech for a few minutes before starting to interpret. The processing can be longer if the speaker forgets to work as a team, which is not rare. The pressure of holding information for that long will undoubtedly overload most people's WM systems. Notepads, therefore, would be an external memory store for interpreters.

Human faces and interpreters' notepads are similar in containing rich information. Although during notetaking, interpreters would be more anxious about recording rather than decoding information, they must interact with the notes to a degree so that units of information and logic can be highlighted in notes. Even if they try not to interact with notes, the eye-mind interaction will force some information into their cognitive system when notetaking is monitored. Interpreters must disengage from notepads when the load becomes high because active listening must not stop. Therefore, GA would be expected during CI.

However, the function of GA could be seen differently. In conversational settings, GA indicates high cognitive load and enhances performance. In CI, the role of GA can still be a cognitive load indicator. However, to be exact, GA would be indicating cognitive crisis. As shown in the results, GA was associated with performance deterioration. Given the gravity of notetaking, it would be warranted to argue that interpreters would endeavour to stay engaged with the task. Therefore, when they conduct GA during notetaking, the cognitive cost associated with that GA can be exceptionally high. Therefore, in CI settings, GA perhaps would be better seen as an indicator of cognitive crisis or even task failure.

6.6 Implications for Practice

This thesis has made implications in several regards. This thesis has bridged Interpreting Studies and Cognitive Psychology. In interpreting studies, scholars on interpreters' cognitive load have traditionally explored fixations and saccades. This thesis has broken new ground for load-related topics to be investigated with a new cognitive load indicator. Also, this thesis contributed to the ongoing discussion about CI notetaking. This thesis would shed new light on the instruction of notetaking for interpreting training. Instructors would be able to better deduce if the trainee interpreters are experiencing cognitive challenges during notetaking and perhaps could even alter module designs to better cultivate information processing and notetaking skills.

For cognitive psychologists, this thesis has paved a new avenue for GA to be explored in interpreting. Also, the thesis has enriched the interpretation of GA – besides indicating and regulating cognitive load, GA can signal cognitive crisis.

6.7 Strengths and Limitations

This thesis has the privilege of enjoying a few strengths. To begin with, to the author's best knowledge, this thesis constitutes one of the earliest efforts in exploring GA in CI active listening and notetaking, and therefore the first and foremost strength would be the courage and originality that this thesis has shown. Also, this thesis has made original contributions to knowledge by introducing GA as a reliable cognitive load indicator and nominating GA as an indicator for cognitive crisis in the context of CI. Third, from the perspective of methodology, this thesis constructed a corpus to study GA in interpreting settings, demonstrating academic rigour and commitment. By listing the exhaustive procedure of constructing and coding the corpus, this thesis would offer significant referential value to future studies. As mentioned before, this thesis coincided with the Covid-19 pandemic. Therefore, being able to swiftly redesign the study and using smartphones to track participants' eye movements showed admirable creativity and resilience.

Like all studies, this thesis is not free from limitations. Perhaps the most significant limitation would be that eye trackers were not used, which would mean that some data known to indicate cognitive load, such as pupillary dilation, were unavailable. However, the pandemic was beyond the author's control. Another limitation would be the small sample size in both studies. As a quantitative study, it would be better if more participants were recruited. However, the number of qualified candidates was small, and therefore the limitation would be acceptable. Last but not least, in the experimental study, participants' WM spans were not tested, which might have cost some robustness in the discussion. But this issue was realised after analysis was completed, by which time the participants have left the programme.

6.8 Directions for future research

Although this thesis has broken the ground GA to be explored in interpreting settings, given the intricacy of both topics, the thesis has only scratched the surface of the water. Findings and limitations in this thesis would be the general direction for future interests. To begin with, future studies could replicate or mimic the experimental study and use eye-tracking technology to see if the data would still lead to a similar conclusion. Second, future efforts in exploring GA in CI would need to consider individual differences in WM at the stage of research design. From the perspective of this thesis, it would be deemed necessary for a future researcher to test participants' WM span so that the data interpretation would be more robust.

Additionally, future studies could also consider using more up-to-date notetaking combinations, such as iPad and Apple Pencil, as it is becoming a trend for interpreters to take notes in such a fashion.

References

Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects of notes, lecture rate, and informational density. *Journal of educational psychology, 67*(3), 439.

Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to eye contact in the West and East: Autonomic responses and evaluative ratings. *PloS one, 8*(3), e59312.

Alamargot, D., Dansac, C., Chesnet, D., & Fayol, M. (2007). Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production. *Studies in writing*, *20*, 13.

Albl-Mikasa, M. (2008). (Non-) Sense in note-taking for consecutive interpreting. *Interpreting*, *10*(2), 197-231.

Albl-Mikasa, M. (2016). Notation Language and Notation Text: A cognitive-linguistic model of consecutive interpreting. *Consecutive notetaking and interpreter training*, 71-117.

Albl-Mikasa, M. (2017). ELF and translation/interpreting. In *The Routledge handbook of English as a lingua franca (pp. 369-383)*. Routledge.

Albl-Mikasa, M. (2019). Note-taking. In *Routledge encyclopedia of translation studies* (pp. 380-386): Routledge.

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In *Teaching Translation and Interpreting 2* (pp. 199). John Benjamins.

Alessandrini, M. S. (1990). Translating numbers in consecutive interpretation: An experimental study.

Al-Jarf, R. (2018). Effect of background knowledge on auditory comprehension in interpreting courses. *Interpretation of Meaning Across Discourses, Brno*.

Alves, R. A., Castro, S. L., de Sousa, L., & Strömqvist, S. (2007). Influence of typing skill on pause-execution cycles in written composition. *Writing and cognition: Research and applications*.

Amano, S. Some observations on neighborhood statistics of spoken English words.

Anderson, A. C., Anderson, A., & Lynch, T. (1988). *Listening.* Oxford University Press.

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220*(4592), 25-30.

Anderson, S., & Kraus, N. (2010). Sensory-Cognitive Interaction in the Neural Encoding of Speech in Noise: A Review. *Journal of the American Academy of Audiology*, *21*(9), 575-585.

Argyle, M. (1996). *The social psychology of leisure.*

Argyle, M. (2013). *Bodily communication.* Routledge.

Argyle, M. (2017). *Social interaction: process and products.* Routledge.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze.*

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry,* 289-304.

Argyle, M., Lefebvre, L., & Cook, M. (1974). The meaning of five patterns of gaze. *European journal of social psychology, 4*(2), 125-136.

Arizpe, J., Walsh, V., Yovel, G. & Baker, C. I. The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. *Vision Research* (2016).

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation, 2*(4), 89-195.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific american, 225*(2), 82-91.

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and linguistic computing, 7*(1), 1-16.

Baddeley, A., & Lewis, V. (1981). Reading: The Inner Voice, the Inner Ear, and the Inner Eye. *Interactive processes in reading*, 107.

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

Baddeley, A. (2010). Working memory. *Current biology, 20*(4), R136-R140.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology,* 63, 1-29.

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113*(4), 518.

Baddeley, A. D. (1982). Domains of recollection. *Psychological review, 89*(6), 708.

Baddeley, A. D. (2002). Is Working Memory Still Working*? European Psychologist, 7*(2), 85-97.

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex,* 112, 91-106.

Barbier, M., & Piolat, A. (2005). L1 and L2 cognitive effort of notetaking and writing. *Proceedings of the* Barlow, M. C., & Miner, L. (1969). Temporal reliability of length-complexity index. *Journal of Communication Disorders, 2*(3), 241-251.

Barnett, J. E., Di Vesta, F. J., & Rogozinski, J. T. (1981). What is learned in note taking? *Journal of educational psychology, 73*(2), 181.

Baron-Cohen, S., Wheelwright, S., Jolliffe, & Therese. (1997). Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*(3), 311-331.

Bayliss, A. P., Paul, M. A., Cannon, P. R., & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review, 13*(6), 1061-1066.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision, 9*(10), 7-7.

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of vision, 11*(10), 6-6.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*(5890), 851-854.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology, 2*(142-162).

Biber, D. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure* (Vol. 28): John Benjamins Publishing.

Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of Speech. The eighth international conference on Language Resources and Evaluation,

Bigi, B. (2014). The SPPAS participation to Evalita 2014. *The SPPAS participation to Evalita 2014,* 127-130.

Bigi, B., & Hirst, D. (2012). SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. *Speech Prosody,*

Bizzocchi, A. L. (2017). How many phonemes does the English language have? *Int. J. Studies English Lang. Lit, 5*(10), 36-46.

Besner, D., Davies, J., & Daniels, S. (1981). Reading for meaning: The effects of concurrent articulation. *The Quarterly Journal of Experimental Psychology, 33*(4), 415-437.

Bohannon, L. S., Herbert, A. M., Pelz, J. B., & Rantanen, E. M. (2013). Eye contact and video-mediated communication: A review. *Displays, 34*(2), 177-185.

Bong, C. L., Fraser, K., & Oriot, D. (2016). Cognitive load and stress in simulation. *In Comprehensive healthcare simulation: pediatrics* (pp. 3-17). Springer.

Brady, S. (1986). Short-term memory, phonological processing, and reading ability. *Annals of dyslexia, 36*(1), 138-153.

Bretzing, B. H., & Kulhavy, R. W. (1981). Note-taking and passage style. *Journal of educational psychology, 73*(2), 242.

Brown, G. D. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behaviour Research Methods, Instruments, & Computers, 16*(6), 502-532.

Bruce, V. (2017). *Recognising faces* (Vol. 3). Routledge.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British journal of psychology, 77*(3), 305-327.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*(1-2), 115-132.

Brünken, R., Seufert, T., & Paas, F. (2010). *Measuring cognitive load*.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology.*

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism – language and cognition, 20*(3), 530-548.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current directions in psychological science*, *27*(1), 45-50.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In *Eye guidance in reading and scene perception* (pp. 125-147). Elsevier.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Bovair, S., & Kieras, D. E. (1981). *A guide to propositional analysis for research on technical prose*. Retrieved from

Braun, S. (1996). Inacuracy for numerals in simultaneous interpretation: neorolinguistic and neuropsychlogical perspectives. In: *Edizioni LINT Trieste*.

Brennan, J. R., & Pylkkänen, L. (2017). MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive science*, 41, 1515-1531.

Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of verbal learning and verbal behavior, 21*(4), 421-436.

Brugman, H., Russel, A., & Nijmegen, X. (2004). *Annotating Multi-media/Multi-modal Resources with ELAN.* Paper presented at the LREC.

Caplan, D., & Waters, G. (1998). *Verbal working memory and sentence comprehension.*

Caplan, D., & Waters, G. (2001). Working memory and syntactic processing in sentence. comprehension. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society, 8*(1), 10-24.

Capozzi, F., Bayliss, A. P., Elena, M. R., & Becchio, C. (2015). One is not enough: Group size modulates social gaze-induced object desirability effects. *Psychonomic Bulletin & Review, 22*(3), 850-855.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological review, 82*(1), 45.

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences*, *98*(9), 5363-5367.

Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL quarterly, 17*(4), 553-573.

Carter, J. F., & Van Matre, N. H. (1975). Note taking versus note having. *Journal of educational psychology, 67*(6), 900.

Celce-Murcia, M. (1995a). Discourse analysis and the teaching of listening. *Principle and practice in applied linguistics: Studies in honor of HG Widdowson*, 363377.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995b). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied linguistics, 6*(2), 5-35.

Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting, 16*(2), 139-168.

Chan, T. M. V. (2020). *Dynamics of Auditory Reflective Attention in Speech-in-noise Perception*

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological review, 113*(2), 234.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive development*, *15*(4), 481-498.

Chang, Y.-j., & Huang, H.-T. (2015). Exploring TED talks as a pedagogical resource for oral presentations: A corpus-based move analysis. *英語教學期刊, 39*(4), 29-62.

Chaudron, C., Loschky, L., & Cook, J. (1994). Second language listening comprehension and lecture note-taking. *Academic listening: Research perspectives*, 75-92.

Chen, I., & Chang, C. C. (2009). Cognitive load theory: *An empirical study of anxiety and task performance in language learning.*

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. Perspectives, 25(4), 640-657.

Chen, S. (2021). The process and product of note-taking and consecutive interpreting: empirical data from professionals and students. *Perspectives*, 1-17.

Chen, S. (2016). Note taking in consecutive interpreting: A review with special focus on Chinese and English literature. *The Journal of Specialised Translation, 26*(1), 151-171.

Chen, S. (2017). Note-taking in consecutive interpreting: New data from pen recording. *Translation & Interpreting, 9*(1), 4-23.

Chen, S. (2018). A pen-eye-voice approach towards the process of note-taking and consecutive interpreting: An experimental design. *International Journal of Comparative Literature and Translation Studies, 6*(2), 1-8.

Choi, H.-H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational psychology review, 26*(2), 225-244.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition, 13*(1991), 127-149.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20*(3), 231-243.

Clerehan, R. (1995). Taking it down: Notetaking practices of L1 and L2 students. *English for specific purposes, 14*(2), 137-155.

Coffey, E. B., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing research*, 352, 49-69.

Cohen, J. D. (1999). 11 A Biologically Based Computational Model of Working Memory. *Models of working memory: Mechanisms of active maintenance and executive control*, 375.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Routledge.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210*(4466), 207-210.

Conrad, L. (1989). The effects of time-compressed speech on native and EFL listening comprehension. *Studies in Second Language Acquisition, 11*(1), 1-16.

Cook, M., Smith, J., & Lalljee, M. G. (1974). Filled pauses and syntactic complexity. *Language and Speech, 17*(1), 11-16.

Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinãngo, M., Balogh, J., & Grossman, M. (2002). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human brain mapping, 15*(2), 80-94.

Coutrot, A., Binetti, N., Harrison, C., Mareschal, I. & Johnston, A. Face exploration dynamics differentiate men and women. *Journal of Vision* 16, 1–19 (2016).

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169, 323-338.

Cox, E., & Salaets, H. (2019). Accuracy: Omissions in consecutive versus simultaneous interpreting. *Student Work*, *64*, 46.

Crawford, C. (1925). Some experimental studies of the results of college note-taking. *The Journal of educational research, 12*(5), 379-386.

Creswell, J. W. (2021). *A concise introduction to mixed methods research*: SAGE publications.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behaviour, 11*(6), 671-684.

Crain, S., & Shankweiler, D. (1988). Syntactic complexity and reading acquisition. *Linguistic complexity and text comprehension: Readability issues reconsidered*, 167-192.

Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring load on working memory: the use of heart rate as a means of measuring chemistry students' cognitive load. *Journal of Chemical Education, 91*(5), 641-647.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition*, 10, 65-70.

Cutler, A., & Clifton Jr, C. (1999). Comprehending spoken language: A blueprint of the listener. In *The neurocognition of language* (pp. 123-166). Oxford University Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive psychology, 24*(3), 381-410.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113.

Dadds, M. R., Allen, J. L., Oliver, B. R., Faulkner, N., Legge, K., Moul, C., Woolgar, M., & Scott, S. (2012). Love, eye contact and the developmental origins of empathy v. psychopathy. *The British Journal of Psychiatry, 200*(3), 191-196.

Dahan, D., & Brent, M. R. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology: General, 128*(2), 165.

Dahan, D., & Ferreira, F. (2019). Language Comprehension: Insights from Research on Spoken Language. *Human Language: From Genes and Brains to Behaviour*, 21.

Dam, H. V. (2004). Interpreters' notes: On the choice of language. *Interpreting, 6*(1), 3-17.

Dam, H. V. (2010). Consecutive interpreting. *Handbook of translation studies,* 1, 75-79.

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

Dawson, K., & Whalen, D. (2017). Influences on lip-reading ability: Aging, sensory, and cognitive functions. *Age*, 25, 59-75.

de-Dios-Flores, I. (2019). Processing sentences with multiple negations: Grammatical structures that are perceived as unacceptable. *Frontiers in Psychology, 10*, 2346.

Di Vesta, F. J., & Gray, G. S. (1972). Listening and note taking. *Journal of educational psychology, 63*(1), 8.

Di Vesta, F. J., & Gray, G. S. (1973). Listening and note taking: II. Immediate and delayed recall as functions of variations in thematic continuity, note taking, and length of listening-review intervals. *Journal of educational psychology, 64*(3), 278.

Diriker, E. (2015). Conference interpreting. In *The Routledge Handbook of Interpreting* (pp. 183-197). Routledge.

Doherty-Sneddon, G. (2004). Don't look now... I'm trying to think. *The Psychologist, 17*(2), 82-85.

Doherty-Sneddon, G., Bonner, L., & Bruce, V. (2001). Cognitive demands of face monitoring: Evidence for visuospatial overload [journal article]. *Memory & Cognition, 29*(7), 909-919. https://doi.org/10.3758/bf03195753

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental psychology, 38*(3), 438.

Doherty-Sneddon, G., & Kent, G. (1996). Visual signals and the communication abilities of children. *Journal of Child Psychology and Psychiatry, 37*(8), 949-959.

Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? [journal article]. *Memory & Cognition, 33*(4), 727-733. https://doi.org/10.3758/bf03195338

Doherty-Sneddon, G., Phelps, F. G., & Calderwood, L. (2009). Gaze aversion during children's transient knowledge and learning. *Cognition and Instruction*, *27*(3), 225-238.

Doherty-Sneddon, G., Riby, D. M., Calderwood, L., & Ainsworth, L. (2009). Stuck on you: Face-to-face arousal and gaze aversion in Williams syndrome. *Cognitive Neuropsychiatry, 14*(6), 510-523. https://doi.org/10.1080/13546800903043336

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

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology, 16*(1-2), 79-103.

Duchowski, A. T., & Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Dunkel, P., Mishra, S., & Berliner, D. (1989). Effects of note taking, memory, and language proficiency on lecture learning for native and non-native speakers of English. *TESOL quarterly*, *23*(3), 543-549.

Edelmann, R. J., & Hampson, S. E. (1979). Changes in non-verbal behaviour during embarrassment. *British Journal of Social and Clinical Psychology, 18*(4), 385-390.

Edwards, S. G., & Bayliss, A. P. (2019). Seeing eye-to-eye: Social gaze interactions influence gaze direction identification. *Attention, Perception, & Psychophysics, 81*(8), 2755-2765.

Ehrlichman, H. (1981). From gaze aversion to eye-movement suppression: An investigation of the cognitive interference explanation of gaze patterns during conversation. *British Journal of Social Psychology, 20*(4), 233-241.

Einstein, G. O., Morris, J., & Smith, S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of educational psychology, 77*(5), 522.

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Ishk.

Ellis, N. C. (2016). Online processing of verb–argument constructions: Lexical decision and meaningfulness. *Language and Cognition, 8*(3), 391-420.

Ellis, N. C. (2019). Usage-based theories of Construction Grammar: Triangulating corpus linguistics and psycholinguistics. In *Using corpus methods to triangulate linguistic analysis* (pp. 239-267): Routledge.

Ellsworth, P. C. (1975). Direct gaze as a social stimulus: The example of aggression. In *Nonverbal communication of aggression* (pp. 53-75). Springer.

El Yagoubi, R., Chiarelli, V., Mondini, S., Perrone, G., Danieli, M., & Semenza, C. (2008). Neural correlates of Italian nominal compounds and potential impact of headedness effect: An ERP study. *Cognitive Neuropsychology, 25*(4), 559-581.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current directions in psychological science, 11*(1), 19-23.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190-193.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309.

Exline, R., Gray, D., & Schuette, D. (1965). Visual behaviour in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology, 1*(3), 201.

Exline, R., Gray, D., & Schuette, D. (1965). Visual behaviour in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology, 1*(3), 201.

Exline, R. V., Ellyson, S. L., & Long, B. (1975). Visual behaviour as an aspect of power role relationships. In *Nonverbal communication of aggression* (pp. 21-52). Springer.

Exline, R. V., & Winters, L. C. (1965). Affective relations and mutual glances in dyads. *Affect, cognition and personality*. New York: Springer, 965.

Fallon, M., Peelle, J. E., & Wingfield, A. (2006). Spoken sentence processing in young and older adults modulated by task demands: evidence from self-paced listening. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 61*(1), P10-P17.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602-9605.

Fedorenko, E., Gibson, E., & Rohde, D. (2004). Verbal working memory in sentence comprehension. Proceedings of the Annual Meeting of the Cognitive Science Society,

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541-553.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(3), 453-489.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210-233.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science, 11*(1), 11-15.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology, 47*(2), 164-203.

Ferreira, F. (2021). In defense of the passive voice. *American Psychologist, 76*(1), 145.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of experimental psychology: learning, memory, and cognition, 22*(2), 324.

Ferrone, L., & Zanzotto, F. M. (2020). Symbolic, Distributed, and Distributional Representations for Natural Language Processing in the Era of Deep Learning: A Survey. *Frontiers in Robotics and AI, 6*, 153.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: sage.

Field, J. (2009). *Listening in the Language Classroom*. Cambridge University Press.

Field, J. (2019). *Second language listening: Current ideas, current issues*. In. Cambridge University Press.

Fisher, J. L., & Harris, M. B. (1973). Effect of note taking and review on recall. *Journal of educational psychology, 65*(3), 321.

Fisher, J. L., & Harris, M. B. (1974). Effect of note-taking preference and type of notes taken on memory. *Psychological Reports, 35*(1), 384-386.

Fiorentino, R., Naito-Billen, Y., Bost, J., & Fund-Reznicek, E. (2014). Electrophysiological evidence for the morpheme-based combinatoric processing of English compounds. *Cognitive Neuropsychology, 31*(1-2), 123-146.

Forster, K. I. (1981). Frequency blocking and lexical access: One mental lexicon or two? *Journal of verbal learning and verbal behaviour, 20*(2), 190-203.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: learning, memory, and cognition, 10*(4), 680.

Foulin, J.-N. (1995). Pauses et débits: les indicateurs temporels de la production écrite. *L'année psychologique, 95*(3), 483-504.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*.

Frazier, L. (2013). Syntax in sentence processing. *Sentence processing, 446*.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Gafiyatova, E., & Pomortseva, N. (2016). The role of background knowledge in building the translating/interpreting competence of the linguist. *Indian Journal of Science and Technology, 9*(16), 2-11.

Gee, K. L. (2011). *The impact of instructor-provided lecture notes and learning interventions on student note taking and generative processing.*

Gerald, B. (2018). A brief review of independent, dependent and one sample t-test. *International journal of applied mathematics and theoretical physics, 4*(2), 50-54.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and individual differences, 102*, 74-78.

Gile, D. (1997). Conference interpreting as a cognitive management problem. *Applied Psychology-London-Sage-*, 3, 196-214.

Gile, D. (2001). Consecutive vs. simultaneous: Which is more accurate. *Interpretation Studies, 1*(1), 8-20.

Gile, D., & Chai, M. J. (2009). *Basic concepts and models for interpreter and translator training* (Vol. 8): John Benjamins Publishing Company Amsterdam/Philadelphia.

Gillies, A. (2017). *Note-taking for consecutive interpreting: A short course*. Taylor & Francis.

Gillies, A. (2019). *Consecutive interpreting: A short course*: Routledge.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 381-398.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*(4), 651-658.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research, 1*(1), 19-33.

Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, 136, 359-364.

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction, 9*(3), 201-219.

Goodman, G. O., McClelland, J. L., & Gibbs, R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition, 9*(6), 580-586.

Gorgoraptis, N., Catalao, R. F., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *Journal of Neuroscience*, *31*(23), 8502-8511.

Gordon, P. C., & Lowder, M. W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Language and Linguistics Compass, 6*(7), 403-415.

Gough, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *Journal of verbal learning and verbal behavior, 5*(5), 492-496.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of experimental psychology: learning, memory, and cognition, 11*(3), 501.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*(2), 228-244.

Greene, E. B. (1928). The relative effectiveness of lecture and individual reading as methods of college teaching. *Genetic Psychology Monographs*.

Grynszpan, O., Martin, J.-C., & Fossati, P. (2017). Gaze leading is associated with liking. *Acta psychologica*, 173, 66-72.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. Proceedings of the 12th ACM international conference on Ubiquitous computing,

Hale, G. A., & Courtney, R. (1994). The effects of note-taking on listening comprehension in the Test of English as a Foreign Language. *Language testing, 11*(1), 29-47.

Hale, S., & Napier, J. (2013). *Research methods in interpreting: A practical resource*. A&C Black.

Hartley, J., & Davies, I. K. (1978). Note-taking: A critical review. *Programmed Learning and Educational Technology, 15*(3), 207-224.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation?: An examination of negated metaphors. *Journal of Pragmatics, 38*(7), 1015-1032.

Heinrich, A., Schneider, B. A., & Craik, F. I. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Quarterly journal of experimental psychology, 61*(5), 735-751.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences, 7*(11), 498-504.

Herbig, N., Düwel, T., Helali, M., Eckhart, L., Schuck, P., Choudhury, S., & Krüger, A. (2020, July). Investigating multi-modal measures for cognitive load detection in e-learning. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 88-97).

Herrero, E. C. (2017). A critical review of listening comprehension in interpreter training: The case of Spanish translation and interpreting degrees. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*(28), 7-22.

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology: Human Perception and Performance, 41*(2), 385.

Howe, M. J. (1970). Using students' notes to examine the role of the individual learner in acquiring meaningful subject matter. *The Journal of educational research, 64*(2), 61-63.

Huette, S., Winter, B., Matlock, T., Ardell, D. H., & Spivey, M. (2014). Eye movements during listening reveal spontaneous grammatical processing. *Frontiers in psychology*, 5, 410.

Hutton, S. (2019). Eye tracking methodology. In *Eye Movement Research* (pp. 277-308): Springer.

Isaacs, G. (1989). Lecture note-taking, learning and recall. *Medical Teacher, 11*(3-4), 295-302.

Ilg, G., & Lambert, S. (1996). Teaching consecutive interpreting. *Interpreting, 1*(1), 69-99.

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field methods, 18*(1), 3-20.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of experimental psychology: General, 118*(2), 115.

Jansen, R. S., Lakens, D., & IJsselsteijn, W. A. (2017). An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review*, 22, 223-233.

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: towards a model of real-world reading. *Journal of Computer Assisted Learning, 33*(3), 193-201.

Johnstone, B. (2017). *Discourse analysis*: John Wiley & Sons.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioural and neural correlates. *Cerebral Cortex*, *15*(10), 1609-1620.

Jones, R. (2014). *Conference interpreting explained*. Routledge.

Just, M. A., & Carpenter, P. A. (1976). The role of eye-fixation research in cognitive psychology. *Behaviour Research Methods & Instrumentation, 8*(2), 139-143.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review, 87*(4), 329.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review, 99*(1), 122.

Kablan, Z., & Erden, M. (2007). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education, 51*(2), 660-668.

Kamide, Y., Scheepers, C., & Altmann, G. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37-55.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of experimental psychology: learning, memory, and cognition, 26*(2), 336.

Kaiser, M.-L., Albaret, J.-M., & Doudin, P.-A. (2009). Relationship between visual-motor integration, eye-hand coordination, and quality of handwriting. *Journal of Occupational Therapy, Schools, & Early Intervention, 2*(2), 87-95.

Kalia, V., Wilbourn, M. P., & Ghio, K. (2014). Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology*, 26(7), 699-713.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono-and disyllabic words and nonwords. *Frontiers in psychology*, 1, 174.

Kidwell, M., & Zimmerman, D. H. (2007). Joint attention as action. *Journal of Pragmatics, 39*(3), 592-611.

Kiewra, K. A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational psychologist, 20*(1), 23-32.

Kiewra, K. A. (1987). Notetaking and review: The research and its implications. *Instructional Science, 16*(3), 233-249.

Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational psychology review, 1*(2), 147-172.

Kiewra, K. A. (2016). Note taking on trial: A legal application of note-taking research. *Educational Psychology Review, 28*(2), 377-384.

Kiewra, K. A., & Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemporary Educational Psychology, 13*(1), 33-44.

Kiewra, K. A., Benton, S. L., Kim, S.-I., Risch, N., & Christensen, M. (1995). Effects of note-taking format and study technique on recall and relational performance. *Contemporary Educational Psychology, 20*(2), 172-187.

Kiewra, K. A., Benton, S. L., & Lewis, L. B. (1987). Qualitative aspects of notetaking and their relationship with information-processing ability and academic achievement. *Journal of Instructional Psychology, 14*(3), 110.

Kiewra, K. A., DuBois, N. F., Christian, D., & McShane, A. (1988). Providing study notes: Comparison of three types of notes for review. *Journal of educational psychology, 80*(4), 595.

Kiewra, K. A., DuBois, N. F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*(2), 240.

Kiewra, K. A., DuBois, N. F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of educational psychology, 83*(2), 240.

Kiewra, K. A., & Fletcher, H. J. (1984). The relationship between levels of note-taking and achievement. *Human Learning: Journal of Practical Research & Applications*.

Kiewra, K. A., & Frank, B. M. (1988). Encoding and external-storage effects of personal lecture notes, skeletal notes, and detailed notes for field-independent and field-dependent learners. *The Journal of Educational Research, 81*(3), 143-148.

Kiewra, K. A., Mayer, R. E., Christensen, M., Kim, S.-I., & Risch, N. (1991). Effects of repetition on recall and note-taking: Strategies for learning from lectures. *Journal of educational psychology, 83*(1), 120.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29*(2), 303-323.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Kintsch, T. A. v. D. W., & Van Dijk, T. (1978). Cognitive psychology and discourse: Recalling and summarizing stories. *Current trends in textlinguistics*, 61.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review, 85*(5), 363.

Kintsch, W., & Vipond, D. (2014). Reading comprehension and readability in educational practice and psychological theory. *Perspectives on learning and memory*, 329-365.

Kintsch, W., & Walter Kintsch, C. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

Kirschner, P. A. (2002). *Cognitive load theory: Implications of cognitive load theory on the design of learning*. In: Elsevier.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist, 41*(2), 75-86.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin, 100*(1), 78.

Knight, D., Morris, S., Arman, L., Needs, J., & Rees, M. (2021). Processing and (Re) presenting Corpora. In *Building a National Corpus* (pp. 105-135). Springer.

Kobayashi, K. (2006). Combined Effects of Note-Taking/-Reviewing on Learning and the Enhancement through Interventions: A meta-analytic review. *Educational Psychology, 26*(3), 459-477.

Kohn, K., & Albl-Mikasa, M. (2002). Note-taking in consecutive interpreting. On the reconstruction of an individualised language. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*(1).

Korpal, P., & Stachowiak-Szymczak, K. (2020). Combined problem triggers in simultaneous interpreting: Exploring the effect of delivery rate on processing and rendering numbers. *Perspectives, 28*(1), 126-143.

Kriston, A. (2012). The importance of memory training in interpretation. *PCTS Proceedings Professional Communication & Translation Studies, 5*(1), 79-86.

Kuhn, G., Tatler, B. W., & Cole, G. G. (2009). You look where I look! Effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition, 17*(6-7), 925-944.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 952.

Lai, S.-L., & Chen, H.-J. H. (2015). Dictionaries vs concordancers: actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning, 28*(4), 341-363.

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in cognitive sciences*, *4*(2), 50-59.

Lasky, R. E., & Klein, R. E. (1979). The reactions of five-month-old infants to eye contact of the mother and of a stranger. *Merrill-Palmer Quarterly of Behaviour and Development, 25*(3), 163-170.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*(3), 339.

Lederer, M. (1982). Le processus de la traduction simultanée.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review, 74*(6), 431.

Lin, C. Y., & Wang, M. (2018). The use of lexical and sublexical cues in speech segmentation by second language learners of English. *Journal of Second Language Studies, 1*(1), 166-198.

Lin, M.-C. (2006). *The effects of note-taking, memory, and rate of presentation on EFL learners' listening comprehension*. La Sierra University.

Lin, Y. (2022). Eye movements and gestures in simultaneous and consecutive interpreting. *Translation & Interpreting, 14*(1), 214-217.

Lindsay, P. H., & Norman, D. A. (2013). *Human information processing: An introduction to psychology*. Academic press.

Liu, M., & Chiu, Y.-H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting, 11*(2), 244-266.

Liu, M. H. (2008). *Consecutive Interpreting and Notes*. Vol. 34

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in cognitive sciences, 4*(1), 6-14.

Locke, E. A. (1977). An empirical study of lecture note taking among college students. *The Journal of educational research, 71*(2), 93-99.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*(3), 155-158.

Luo, L., Kiewra, K. A., Flanigan, A. E., & Peteranetz, M. S. (2018). Laptop versus longhand note taking: effects on lecture notes and achievement. *Instructional Science, 46*(6), 947-971.

Lynch, T. (1998). Theoretical perspectives on listening. *Annual review of applied linguistics*, 18, 3-19.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*(1), 50-68.

Malah, Z., & Rashid, S. M. (2015). Contrastive analysis of the segmental phonemes of English and Hausa languages. *International Journal of Languages, Literature and Linguistics, 1*(2), 106-112.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1-71.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology Section A, 45*(1), 73-87.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive psychology, 10*(1), 29-63.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology, 8*(4), 506.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 960.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477.

Matysek, H. (1989). Handbuch der Notizentechnik für Dolmetscher: Ein Weg zur sprachunabhängigen Notation. Heidelberg: J. Groos.

Mazzei, C. (2007). *Note-Taking for Consecutive Interpreting*. In.

McAllister, R. (2000). Perceptual foreign accent and its relevance for simultaneous interpreting. *Benjamins Translation Library, 40*, 45-64.

McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze display when thinking depends on culture and context. *Journal of Cross-Cultural Psychology, 39*(6), 716-729.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology, 18*(1), 1-86.

McQueen, J. M., & Cutler, A. (1992). *Words within words: Lexical statistics and lexical access*. The Second International Conference on Spoken Language Processing,

McQueen, J. M., & Cutler, A. (1998). *Morphology in word recognition*. Blackwell.

Mehrabian, A. (2017). *Nonverbal communication*. Routledge.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of verbal learning and verbal behavior, 2*(4), 346-351.

Mellinger, C. D. (2017). Equal access to the courts in translation: a corpus-driven study on translation shifts in waivers of counsel. *Perspectives, 25*(2), 308-322.

Meyer, B. (2008). Interpreting proper names: Different interventions in simultaneous and consecutive interpreting. *Trans-kom, 1*(1), 105-122.

Mitsuda, T., & Glaholt, M. G. (2014). Gaze bias during visual preference judgements: Effects of stimulus category and decision instructions. *Visual Cognition, 22*(1), 11-29.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive neuropsychology, 11*(6), 671-717.

Modigliani, A. (1971). Embarrassment, facework, and eye contact: Testing a theory of embarrassment. *Journal of Personality and Social Psychology, 17*(1), 15.

Monaghan, P., Chang, Y.-N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language*, 93, 1-21.

Moody, B. (2011). What is a faithful interpretation? *Journal of Interpretation, 21*(1), 4.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of experimental psychology: learning, memory, and cognition, 21*(1), 116.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 667.

Morsbach, H. (1973). Aspects of nonverbal communication in Japan. *Journal of Nervous and Mental Disease*.

Moser, D. C., Fridriksson, J., & Healy, E. W. (2007). Sentence comprehension and general working memory. *Clinical linguistics & phonetics, 21*(2), 147-156.

Moser-Mercer, B. (2000). Simultaneous interpreting: Cognitive potential and limitations. *Interpreting, 5*(2), 83-94.

Murphy, D. R., Craik, F. I., Li, K. Z., & Schneider, B. A. (2000). Comparing the effects of aging and background noise on short-term memory performance. *Psychology and aging, 15*(2), 323.

Mustafić, A. (2020). *The English passive: students' comprehension and preferences.* University of Zagreb. University of Zagreb, Faculty of Humanities and Social

Mutter, S. A., & Hashtroudi, S. (1987). Cognitive effort and the word frequency effect in recognition and lexical decision. *The American journal of psychology*, 93-116.

Nelson, O. A. (2011). *Acoustic properties of coda liquids in Californian English.* Paper presented at the Proceedings of Meetings on Acoustics 162ASA.

Neville, D. A., Raaijmakers, J. G., & van Maanen, L. (2019). Modulation of the word frequency effect in recognition memory after an unrelated lexical decision task. *Journal of Memory and Language*, 108, 104026.

Norman, D. A., & Shallice, T. (1986). *Attention to action.* In *Consciousness and self-regulation* (pp. 1-18). Springer.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of experimental psychology: learning, memory, and cognition, 21*(5), 1209.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive psychology, 34*(3), 191-243.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied linguistics, 30*(4), 555-578.

Nurmsoo, E., Einav, S., & Hood, B. M. (2012). Best friends: Children use mutual gaze to identify friendships in others. *Developmental science, 15*(3), 417-425.

Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology*, *21*(5), 758-785.

Orenes, I., Moxey, L., Scheepers, C., & Santamaría, C. (2016). Negation in context: Evidence from the visual world paradigm. *Quarterly Journal of Experimental Psychology, 69*(6), 1082-1092.

Orlando, M. (2010). *Digital pen technology and consecutive interpreting: another dimension in notetaking training and assessment*.

Osterhout, L. (1994). Event-related brain potentials as tools for comprehending language comprehension. *Perspectives on sentence processing*, 15-44.

Otake, T., & Cutler, A. (2013). Lexical selection in action: Evidence from spontaneous punning. *Language and Speech, 56*(4), 555-573.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist, 38*(1), 1-4.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist*, *38*(1), 63-71.

Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human factors, 35*(4), 737-743.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review*, *6*(4), 351-371.

Peper, R. J., & Mayer, R. E. (1978). Note taking as a generative activity. *Journal of educational psychology, 70*(4), 514.

Peper, R. J., & Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *Journal* of educational psychology, 78(1), 34.

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of vision*, 16(7), 12-12.

Peverly, S. T., Vekaria, P. C., Reddington, L. A., Sumowski, J. F., Johnson, K. R., & Ramsay, C. M. (2013). The relationship of handwriting speed, working memory, language comprehension and outlines to lecture note-taking and test-taking among college students. *Applied cognitive psychology*, *27*(1), 115-126.

Peverly, S. T., & Wolf, A. D. (2019). *Note-taking*.

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of experimental psychology: learning, memory, and cognition*, *27*(1), 139.

Phelan, M. (2001). *TheInterpreter's Resource*: Multilingual Matters.

Phelps, F., Doherty-Sneddon, G., & Warnock, H. (2006). Functional benefits of children’s gaze aversion during questioning. *British journal of developmental psychology*, *24*, 577-588.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological bulletin, 144*(10), 1002.

Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied cognitive psychology, 19*(3), 291-312.

Piolat, A. (2007). Effects of note taking technique and working-memory span on cognitive effort. *G. Rijlaarsdam (Series Ed.), Studies in Writing, & D. Galbraith, M. Torrance, & L. van Waes (Vol. Eds.), Recent developments in writing process research, 1*.

Piolat, A., Barbier, M.-L., & Roussey, J.-Y. (2008). Fluency and Cognitive Effort During First-and Second-Language Notetaking and Writing by Undergraduate Students. *European Psychologist, 13*(2), 114-125.

Piolat, A., Roussey, J.-Y., Olive, T., & Farioli, F. (1996). Charge mentale et mobilisation des processus rédactionnels: examen de la procédure de Kellogg. *Psychologie française, 41*(4), 339-354.

Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge university press.

Pöchhacker, F. (2005). From operation to action: Process-orientation in interpreting studies. *Meta: journal des traducteurs/Meta: Translators' Journal, 50*(2), 682-695.

Pöchhacker, F. (2011). Consecutive interpreting. In *The Oxford Handbook of Translation Studies*.

Pöchhacker, F. (2016). *Introducing interpreting studies*. Routledge.

Pöchhacker, F., & Shlesinger, M. (2002). *The interpreting studies reader*: Psychology Press.

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological review, 127*(1), 1.

Pouw, W., & Dixon, J. A. (2019). *Quantifying gesture-speech synchrony*. Paper presented at the the 6th gesture and speech in interaction conference.

Race, E., Burke, K., & Verfaellie, M. (2019). Repetition priming in amnesia: Distinguishing associative learning at different levels of abstraction. *Neuropsychologia*, 122, 98-104.

Radach, R., Kennedy, A., & Rayner, K. (2004). *Eye movements and information processing during reading* (Vol. 11). Psychology Press.

Ran, H. (2021). *Written English Syntactic Complexity Analysis Based on AES System and Big-data Technology*. 2021 International Conference on Aviation Safety and Information Technology,

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191-201.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1188.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence–picture verification. *Cognitive psychology, 40*(4), 261-295.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of verbal learning and verbal behavior, 10*(6), 645-657.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological science, 21*(9), 1300-1310.

Rhodes, G., & Haxby, J. (2011). *Oxford handbook of face perception*. Oxford University Press.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL quarterly, 17*(2), 219-240.

Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology, 3*(2), 136-143.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological science, 12*(1), 94-99.

Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports, 8*(1), 1-9.

Rozas, V. V., & Barcala, M. (2020). Computational tools and spoken corpora design: an ongoing dialogue. Caplletra. *Revista Internacional de Filologia*(69), 221-240.

Russell, D., & Takeda, K. (2015). Consecutive interpreting. *The Routledge handbook of interpreting*. London and New York: Routledge, 96-111.

Sanders, L. D., & Neville, H. J. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research, 43*(6), 1301-1321.

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). *Speech segmentation by native and non-native speakers*. the neural correlates of implicit and explicit memory. *Nature, 392*(6676), 595-598.

Sanz-Torrent, M., Andreu, L., Rodriguez Ferreiro, J., Coll-Florit, M., & Trueswell, J. C. (2017). Auditory word recognition of verbs: Effects of verb argument structure on referent identification. *PloS one, 12*(12), e0188728.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 1.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of experimental psychology: learning, memory, and cognition, 13*(3), 501.

Schelble, J. L., Therriault, D. J., & Miller, M. D. (2012). Classifying retrieval strategies as a function of working memory. *Memory & Cognition, 40*(2), 218-230.

Schmitz, J., Scheel, C. N., Rigon, A., Gross, J. J., & Blechert, J. (2012). You don't like me, do you? Enhanced ERP responses to averted eye gaze in social anxiety. *Biological psychology, 91*(2), 263-269.

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: behaviour, theory, and biological mechanisms. *Cognitive science, 27*(3), 525-559.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational psychology review, 19*(4), 469-508.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories—new models. *Interpreting*, 13(2), 176-204.

Seeber, K. G. (2015). Cognitive load. *Routledge Encyclopedia of Interpreting Studies*. London–New York: Routledge, 60.

Seeber, K. G. (2015). *Eye tracking. Pöchhacker, F. Routledge Encyclopedia of Interpreting Studies*. London/New York: Routledge.

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism, 16*(2), 228-242.

Sekkal, K. (2020). Listening Comprehension for'Business English'Students: Obstacles, Needs, and Strategies. *Needs, and Strategies* (January 30, 2020).

Seleskovitch, D. (1975). Language and memory: A study of note-taking in consecutive interpreting. Trans. By Jacolyn Harmer. In F. Pöchhacker (Ed.), The Interpreting Studies Reader, 120-129.

Seleskovitch, D. (1978). Language and cognition. In Language interpretation and communication (pp. 333-341). Springer, Boston, MA.

Seleskovitch, D., & Lederer, M. (1984). *Interpréter pour traduire*. Didier érudition Paris.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current biology, 18*(9), 668-671.

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition, 12*(1), 127-144.

Setton, R., & Dawrant, A. (2016). *Conference Interpreting–A Complete Course* (Vol. 120). John Benjamins Publishing Company.

Sharon, O. B. (2007). Eye-tracking and translation memory matches. *Perspectives: Studies in translatology, 14*(3), 185-205.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review, 84*(2), 127.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological review, 76*(2), 179.

Shorten, A., & Smith, J. (2017). Mixed methods research: expanding the evidence base. In (Vol. 20, pp. 74-75): Royal College of Nursing.

Sinclair, J. (2005). Corpus and text-basic principles. *Developing linguistic corpora: A guide to good practice*, 92, 1-16.

Singer, M. (1994). *Discourse inference processes*.

Slater, A., & Quinn, P. C. (2001). Face recognition in the newborn infant. *Infant and Child Development: An International Journal of Research and Practice, 10*(1-2), 21-24.

Sloetjes, H. (2017). ELAN. In *The Oxford Handbook of Corpus Phonology*.

Someya, Y. (2016). *Consecutive notetaking and interpreter training*. Taylor & Francis.

Someya, Y. (2016). A propositional representation theory of consecutive notes and notetaking. In *Consecutive notetaking and interpreter training* (pp. 157-200). Routledge.

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human–Computer Interaction*, 38(9), 801-824.

Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and aging, 15*(2), 253.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142): Citeseer.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience, 4*(3), 232-243.

Stockmeyer, N. O. (2009). Using Microsoft Word's readability program. *Michigan Bar Journal*, 88, 46.

Supriya, S., & Handore, S. (2017). *Speech recognition using HTK toolkit for Marathi language*. Paper presented at the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI).

Stachowiak-Szymczak, K. (2019). *Eye movements and gestures in simultaneous and consecutive interpreting*. Springer.

Stanley, G., & Martin, D. S. (1968). Eye-contact and the recall of material involving competitive and noncompetitive associations. *Psychonomic Science, 13*(6), 337-338.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. *The Oxford handbook of psycholinguistics*, 327, 342.

Stuart, K., Botella, A., & Ferri, I. (2016, November). A corpus-driven approach to sentiment analysis of patient narratives. In *8th International Conference on Corpus Linguistics* (Vol. 1, pp. 381-395)

Sugar, J., & Moser, M. B. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus, 29*(12), 1190-1205.

Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation* (Vol. 56). Studentlitteratur.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning, 3*(2), 19-30.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review, 22*(2), 123-138.

Sweller, J. (2011). *Cognitive load theory*. *In Psychology of learning and motivation* (Vol. 55, pp. 37-76). Elsevier.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71-85). Springer, New York, NY.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review, 31*(2), 261-292.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review, 10*(3), 251-296.

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In *Eye Movements* (pp. 443-II). Elsevier.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557-580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). *Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation.*

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In *Handbook of psycholinguistics* (pp. 863-900). Elsevier.

Thiéry, C. (1981). L'enseignement de la prise de notes en interprétation consécutive: un faux problème? L'enseignement de la traduction et de l'interprétation. De la théorie à la pédagogie. *Ottawa: Éditions de l'Université d'Ottawa*, 99-112.

Thiéry, C. (1990). The sense of situation in conference interpreting. *Interpreting: Yesterday, Today, and Tomorrow*, 40-43.

Thorley, C. (2016). Note taking and note reviewing enhance jurors' recall of trial information. *Applied cognitive psychology, 30*(5), 655-663.

Tian, Y., & Breheny, R. (2019). Negation. In *The Oxford handbook of experimental semantics and pragmatics*.

Tian, Y., & Breheny, R. E. (2018). Pragmatics and negative sentence processing.

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience, 31*(5), 683-698.

Tiselius, E., & Sneed, K. (2020). Gaze and eye movement in dialogue interpreting: An eye-tracking study. *Bilingualism: Language and Cognition, 23*(4), 780-787.

Tomkins, S. S., & McCarter, R. (1964). What and where are the primary affects? Some evidence for a theory. *Perceptual and Motor Skills, 18*(1), 119-158.

Townsend, D. J., Bever, T. G., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. MIT Press.

Toyama, T., Kieninger, T., Shafait, F., & Dengel, A. (2012). *Gaze guided object recognition using a head-mounted eye tracker.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. Journal of Memory and Language, 35(3), 454-475.

Tsui, A. B., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. Applied linguistics, 19(4), 432-451.

Tulving, E. (1972). Episodic and semantic memory.

Tulving, E. (1983). *Elements of episodic memory*.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne, 26*(1), 1.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology, 53*(1), 1-25.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America, 126*(1), 367-376.

Ujbányi, T., Kővári, A., Sziládi, G., & Katona, J. (2020). Examination of the eye-hand coordination related to computer mouse movement. *Infocommunications Journal, 12*(1), 26-31.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41*(2), 242-254.

Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse studies, 11*(5), 585-605.

Vakil, E., Wasserman, A., & Tibon, R. (2018). Development of perceptual and conceptual memory in explicit and implicit memory systems. *Journal of Applied Developmental Psychology*, 57, 16-23.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*.

Van Gerven, P., Paas, F., Van Merriënboer, J., & Schmidt, H. (2002). Memory load and task-evoked pupillary responses in aging. Manuscript submitted for publication.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology, 67*(6), 1176-1190.

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review, 17*(2), 147-177.

Van Straaten, I., Holland, R. W., Finkenauer, C., Hollenstein, T., & Engels, R. C. (2010). Gazing behavior during mixed-sex interactions: Sex and attractiveness effects. Archives of Sexual Behavior, 39(5), 1055-1062.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual review of applied linguistics, 24*(1), 3-25.

Vandergrift, L. (2011). Second language listening. *Handbook of research in second language teaching and learning*, 2, 455.

Vecera, S. P., & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition, 2*(1), 59-87.

Viezzi, M. (2013). consecutive interpreting (non-conference settings). *The Routledge handbook of translation studies*, 377.

Vitevitch, M. S., & Luce, P. A. (2016). Phonological Neighborhood Effects in Spoken Word Perception and Production. Annu. Rev. *Linguist*, 2, 7.1-7.20.

Vitevitch, M. S., Siew, C. S., & Castro, N. (2018). Spoken word recognition. In *The Oxford handbook of psycholinguistics* (pp. 31). Oxford University Press.

Volkmar, F. R., & Mayes, L. C. (1990). Gaze behaviour in autism. *Development and Psychopathology, 2*(1), 61-69.

Wang, F. H. (2020). Explicit and implicit memory representations in cross-situational word learning. *Cognition*, 205, 104444.

Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Memory and Language, 9*(4), 398.

metacognitive awareness. *System*, 65, 139-150.

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217*(5132), 972-974.

Waters, G., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review, 10*(1), 88-95.

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition, 24*(3), 342-355.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology Section A, 57*(1), 129-163.

Weeks, J. W., Howell, A. N., & Goldin, P. R. (2013). Gaze avoidance in social anxiety disorder. *Depression and anxiety, 30*(8), 749-756.

Weiskrantz, L., & Warrington, E. K. (1979). Conditioning in amnesic patients. *Neuropsychologia, 17*(2), 187-194.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1-12.

White, L. (2018). Segmentation of speech. *The Oxford handbook of psycholinguistics*, 1.

Wickens, C. D. (1991). Processing resources and attention. *Multiple-task performance*, 1991, 3-34.

Wilson, D., & Sperber, D. (2002). Relevance theory. In: Blackwell.

Wingfield, A., Peelle, J. E., & Grossman, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Aging, Neuropsychology, and Cognition, 10*(4), 310-322.

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *Journal of the American Academy of Audiology, 18*(7), 548-558.

Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The episodic L2 hypothesis. *Journal of experimental psychology: learning, memory, and cognition, 38*(6), 1608.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). *ELAN: A professional framework for multimodality research.* Paper presented at the 5th international conference on language resources and evaluation (LREC 2006).

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition, 12*(1), 31-45.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes, 27*(10), 1489-1524.

Wynne, M., Sinclair, J., Leech, G., Burnard, L., McEnery, A., Xiao, R., & Thompson, P. (2000). Developing linguistic corpora. OTA Guides.

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single-and multi-task environments. *Work & stress, 14*(1), 74-99.

Yeldham, M. (2018). L2 Listening Instruction: More Bottom-up or More Top-down? *Journal of Asia TEFL, 15*(3), 805.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517.

Yoon, S.-Y., & Bhat, S. (2018). A comparison of grammatical proficiency measures in the automated assessment of spontaneous speech. *Speech Communication*, 99, 221-230.

Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (pp. 78-85).

Zhang, W. , & Ke, F. (2008). *Looking at assessment of interpreting from audience's perspective (从口译用户看口译质量评估)*. Learning Foreign Language, 3, 114-118.

Appendix A (Consent Form for the Corpus Study)

Consent Form

My name is Wenbo Guo, and I am currently researching cognition in Consecutive Interpreting at Newcastle University. This form is made to obtain your consent in giving access to the recorded video of you during Final Exam of Consecutive Interpreting 1. The focus of the study is your eye-movements and **not your performance/interpretation**.

Please read each of the sections below and sign your name if you consent to the terms of the study and return the signed form to w.guo7@newcatle.ac.uk

Thank you very much.

- I am aware of the purpose of the study;
- I understand that the researcher will not assess or judge my performance;
- I understand that my footage will be stored securely on a password protected server accessible only by the researcher;
- I understand that I am not entitled to the ownership of the data although I can withdraw my participation, partially or wholly, at any stage of the study, without giving a reason;
- I understand that should the result be published, the researcher is not obliged to inform the participants of the publication;
- I understand that my data will be confidential and that any information in the transcript that may potentially identify me will be masked or anonymised;
- I understand that participation in this study is entirely voluntary and without payment of any kind;

Should you agree to the above-mentioned terms, please tick 'Yes', otherwise please tick 'No'.

☐ Yes ☐ No

Additional Signature:
There might be occasions when playing an edited footage can enhance demonstration for teaching or research purposes. I am fully aware that this obviously contradicts the procedures in place to protect your rights. But it would be much appreciated. Please tick one box to inform me your decision as to whether you allow me to use your footage for teaching and research purposes. If you skip this question, the answer is assumed to be 'No'.

□ Yes □ No

Participant signature... Date ……………

Researcher signature…………Wenbo Guo……..... Date ……16/03/2018…

Appendix B (Consent Form for the Experimental Study)

Consent Form

This form to obtain your consent to participate in an experiment on Consecutive Interpreting as a part of a PhD project. Please read each statement carefully and sign your name in the end if you choose to give your consent.

- I, _______________, voluntarily agree to participate in this experiment.
- I understand that even if I agree to participate now, I can withdraw at any time without having to give any reason or face any consequences of any kind.
- I have had the purpose and nature of the study explained to me in writing, and I have had the opportunity to ask questions about the study.
- I understand that by participating in this experiment, I will conduct a Consecutive Interpreting task.
- I understand that a video camera and a Livescribe Smartpen will be used as experiment tools to record my face, voice, and notes.
- I understand that all information I provide for this study will be treated confidentially.
- I understand that in any report on the results of this research, my identity will remain anonymous. This will be done by changing my name and disguising any details which may reveal my identity.
- I understand that signed consent forms and original video recordings will be retained in a hard drive in his personal PC accessed only by the researcher.
- I understand that under freedom of information legalisation I am entitled to access the information I have provided at any time while it is in storage as specified above.

If the results of the project are to be presented at an academic conference, the researcher might need to briefly show some data to the audience. Would you allow using your video during presentations?

- YES

- NO

Signature of participant

Signature of participant Date

Signature of researcher

Signature of researcher Date

Appendix C (Source Speech for the Corpus Study)

<u>Source Speech in 2013</u>

On Dec. 31, 2012, tax cuts and across-the-board government spending cuts are scheduled to become effective. The idea behind the fiscal cliff was that if the federal government allowed these two plans to go ahead, they would have a detrimental effect on an already shaky economy, sending it back into an official recession as it would cut household incomes, increase unemployment rates and undermine consumer and investor confidence. But at the same time, it was predicted, or hoped I should say, that going over the fiscal cliff would significantly reduce the federal budget deficit.

We all know that the U.S. did not go over the cliff on New Year's Day of 2013. While many let out a sigh of relief when Obama passed the fiscal cliff bill, much remains to be done, and much more remains debatable. Much remains to be done because the metaphorical cliff did not disappear, because public debt is still rising and the growth in GDP is slowing. Much more remains debatable because the consequence of going over a real cliff is not helping the general public see the whole picture. When Ben Bernanke coined the term 'fiscal cliff', I think he had a crucial deadline and a sense of urgency in mind rather than a whole country shutting down in a blink of an eye if the Senate and the House of representatives fail to pass the bill. NOW, let's imagine that the bill was not passed in time and the U.S. went over the cliff. Because a cliff is so steep, by the time I finish this sentence, the DOW JONES Industrial Average would have hit the bottom. However, some analysts in Washington argued that the term 'fiscal hill' or 'fiscal slope' might be more accurate. Here is why. It is real that going off the cliff affects 88% of U.S. tax payers, with an average drop of 6.2% in after-tax income and average increase of tax by 3,500 U.S. dollars a year. Although the hit of tax increase may be felt pretty immediately, the spending cuts would be phased in gradually over a decade. In other words, the effects would be powerful but gradual. Then when I contemplated these effects, a word struck me. It's the word of the year 2010. In Economics, the word 'austerity' describes policies used by governments to reduce budget deficits during adverse economic conditions. These policies can include spending cuts, tax increases, or a mixture of the two—exactly what was going to happen if the U.S. went off the cliff. So, why not be upfront with US citizens that it's going to be an era of austerity however their government tries to shield the essential public spending from aggressive cuts. To their people, what they see isn't a cliff, it's a looming crisis of austerity. It shows a lack of long-term plan in their government. It's clear to us all how austerity has maimed the quality of life in Greece and stifled the UK's economy. What the U.S. government should be obsessed with is therefore not an overnight deficit cut, but stimulating demand, creating jobs, and making their healthcare system work more efficiently. Thank you.

Source Speech in 2014

I often meet individuals who are eager to share their views on science despite the fact that they have never done an experiment. They have never tested an idea for themselves, using their own hands; or seen the results of that test, using their own eyes; and they have never thought carefully about what those results mean for the idea they are testing, using their own brain. To these people 'science' is a monolith, a mystery, and an authority, rather than a method. So dismantling outrageous pseudoscientific claims is an excellent way to learn the basics of science, partly because science is largely about disproving theories, but also because the lack of scientific knowledge among miracle-cure therapists and journalists gives us some very simple ideas to test. Their knowledge of science is rudimentary; they also rely on notions like oxygen, water, and toxins.

Take Aqua Detox for example. The hypothesis from these Spas where you can have your detox footbath is very clear: your body is full of 'toxins', whatever those may be. You put your feet in the bath, the toxins are extracted, and the water goes brown. Is the brown in the water because of the toxins? Or is that merely theatre? One way to test this is to go along and have an Aqua Detox treatment yourself at a health spa, and take your feet out of the bath when the therapist leaves the room. If the water goes brown without your feet in it, then it wasn't your feet or your toxins that did it. The brown colour is in fact the result of electrolysis. But therapists would explain that it's the chlorine coming out of your body, from all the plastic packaging on your food, and all those years bathing in chemical swimming pools. And even in samples taken from a genuine detox footbath, hospital analysts didn't manage to identify any toxin in the brown water, just lots of rusty iron.

The detox phenomenon is interesting because it represents one of the most grandiose innovations of marketers, lifestyle gurus, and alternative therapists – it is the invention of a whole new physiological process. There is however nothing on the detox system in a medical textbook. That burgers and beer can have negative effects on your body is certainly true, for a number of reasons; but the notion that they leave a specific residue, which can be extruded by a specific process, is a marketing invention. Because it has no scientific meaning, detox is much better understood as a cultural product. It deliberately blends useful common sense with outlandish, medicalised fantasy. Next time when you consider spending hundreds of pounds on detox products, remember that your liver, kidneys and colon detox our body for free.

Source Speech in 2015

What I'd like to talk about today is the fact that this morning, about a billion people on Earth didn't know what to put on the table for themselves or for their families. In fact, what we know now is that every 10 seconds we lose a child to hunger. Over the past nine years, I have had the unique and life-changing opportunity to travel with the UN World Food Programme and UNICEF to visit countries that are affected by poverty and hunger, such as Guatemala, Sri Lanka, Cambodia, and Kenya to name a few. Poverty and hunger persist in countries and regions for a variety of reasons, social, political and sometimes geographical. But it is a guarantee that where there is poverty there will be hungry people who struggle everyday to feed themselves and their families.

Hunger is often called, in the humanitarian aid world, a 'silent killer. Hunger kills more people every year than AIDS, malaria and TB combined. It goes without saying that food and water are the most essential things needed to sustain life. When a person does not have access to these very basic human needs, they become desperate, angry, and they can easily lose hope and dignity. Now if we look at the economic imperative here, this isn't just about compassion. Studies show that the cost of malnutrition and hunger is on average six percent, and in some countries up to 11 percent, of GDP a year. So my dream is to take this issue, not just from the compassion argument, to the finance ministers of the world, and say we cannot afford to not invest in the access to adequate, affordable nutrition for all of humanity.

Where do we start? Did you know that we waste food, enough to feed 2 billion people, every year? An estimated 1.3 billion tonnes of food, or roughly 30 percent of global production, is lost or wasted according to the U.N. Food and Agriculture Organization. So in developed countries, we need to stop throwing food away simply because we believe it has gone rotten, or when products don't look picture perfect. In developing regions, spoilage happens during storage or transport, as infrastructure for refrigeration and preservation is often inadequate. I have witnessed that better storage facilities reduced post-harvest waste to less than two percent during a 90-day trial period in a WFP project, and the WFP is now scaling up the programme by taking it to 41,000 farmers and aiming for a 70 percent reduction in post-harvest losses. What I would like to offer here is a challenge. I believe we're living at a time in human history where it's just simply unacceptable that children wake up and don't know where to find a cup of food. No more. No more are we going to accept this. And we want to tell our grandchildren that there was a terrible time in history where up to a third of the children had brains and bodies that were stunted, but that exists no more. Thank you.

<u>Source Speech in 2016</u>

I read an article in The Wall Street JournaI today and I can’t think of a better piece of news to begin my talk with. Two Italian economists compared data on Florentine taxpayers from 1427 against tax data from 2011 and found about 900 surnames still present in Florence. It would appear that the wealthiest families in Florence today are descended from the wealthiest families of Florence nearly

600 years ago. I later found out that descendants of Japan’s samurai remain elites 140 years after their ancestors gave up their swords. Even the communist revolution in China failed to drive social mobility. Why does social class and its immobility persist?

The poignant documentary ‘56UP’ shows just how static our society has remained in the last half a century. According to OECD figures, Britain may have some of the lowest levels of social mobility in the developed world. This finding must imply that we know how to determine our social class, or do we? Am I middle-class because I start to replace the word ‘dinner’ with the word ‘tea’? Does it make any difference to your class calculator when you hear ‘How d’you do?’ versus ‘Pleased to meet you’? As if this is not confusing enough, there is actually a BBC class calculator website that groups people into seven classes. Mind you though, among other things, the calculator takes your music taste and who you hang out with into consideration too. It appears that the class that I most closely match is ‘emergent service worker’ – I haven’t got a clue what it means, and why class matters. It’s difficult to grasp, but you feel it everywhere you go. The cultural critic Richard Hoggart once said that ‘each year we shiftily declare we have buried class, yet each decade the coffin stays empty.’

Do we stand a chance of witnessing social mobility in real terms if the government fails to recognise that inequality angers more people than social immobility does. Yet all the while they have this obsession of getting a small number of younger students from poorer backgrounds to Oxbridge. It shows a worrying lack of appreciation of the achievements and the quality of the teaching provided in modern universities. It implies that even universities are stratified too. It is firmly believed that education is the closest thing we have to a silver bullet when it comes to social mobility, but since this government took power, we have seen major financial barriers erected in the face of those from low and average-income backgrounds, exactly contradicting their strategy of ‘opening doors, breaking barriers’. As someone who has been through class struggle, I am afraid that the class escalator today is steeper and much longer.

Appendix D (Source Speech in the Experimental Study)

Easy Speech

Chunk 1:

A few years ago developed countries made a donation. They gave money to the poorest countries . The goal was to solve one of the world's biggest problems. That is hunger .Many NGOs volunteered, too. They helped the poor countries to receive and manage the funds. I work for one of the NGOs. The job is very physically demanding. I live in the UK. But my clients are mainly in Africa and Asia. I have to adapt to their working hours. Regular sleeping hours to me is a luxury. But there is a privilege from my job. That is, I travel a lot. My job does not usually take me to rich countries. It sends me to the opposite. To understand poverty, you have to be there. I have been to many poor countries. I witnessed the daily struggles for food and medicine. I saw people fight to death for clean water. Take Cambodia for example. I lived in Cambodia between two thousand and six and two thousand and nine. Cambodia has been one of the poorest countries. According to World Bank, Cambodia had a poverty rate of 47% in 2007. That is a really big number. I have more figures for you. 10 years ago, forty percent of the population did not have access to life saving medicines. The average life expectancy was sixty for men and sixty two for women; Over five percent of women died during labour; Nearly nine percent of Cambodian children died before age of five; Most people could only have one meal each day; Clean water was a dream for most of them; Primary education was next to none. I only lived in Cambodia for three years. It was a short period of time. But I witnessed some significant transition. With the donation and other forms of help, middle Cambodia accomplished great achievements. Today, most people in middle Cambodia still live in rural areas. But eighteen percent of the local population managed to escape poverty. Modern agricultural machines were donated to the area. A large number of local farmers learnt to use them. Better crops were also brought to the area. So they produced much more food. Almost all adults can have two meals a day. Children can even have three meals. NGOs helped to build hospitals. Most women can give birth in a clean and safe environment. Doctors and health workers are helping mothers to nurture their babies. Free vaccinations for life threatening and infectious diseases are becoming available. Local authorities are also working with NGOs to promote education. Primary schools and vocational centres have been set up across the area. All courses are free and taught by international volunteers. These improvements transformed the area. Millions of people are living a better life. There is still a long way to go. But people now have hopes. And we are all trying hard to give them more hopes.

Chunk 2

In 1990, 36% of the world's population lived in poverty. That was nearly two billion people.
Most of them lived under extreme poverty at that time. In other words, they lived under less than 50 pence a day. A bottle of coca cola would cost you thirty five pence in the UK back then. The world

has been fighting poverty in many ways. Owing to the hard work, the poverty rate dropped to ten percent today. This is a great achievement.

The world is indeed in a better place. But ten percent of the population is still a huge number. Over seven hundred million people are living on less than one pound a day.

That is beyond tragic. Poverty does not just mean no money. It hurts human race several ways. Needless to say, poverty damage our health. In the poorest areas, people suffer from bad living conditions. Most of them don't have windows on the wall. The shortage of fresh air can easily cause lung infections. Toilets are also a luxury. The Lack of proper toilets leads to the spread of life threatening illness. The fragile houses are easily ruined by rain, storms, and earthquakes. Some of the poorest have to live in a tent or under a shelter. Anything they own is at the risk of theft. They could be attacked or even killed by wild animals. Children under poverty sometimes do not get education. They very rarely realise their potential. The poorest population are more vulnerable than the rest of us. We are in the COVID nineteen pandemic now. It is hitting the poorest people particularly hard. Many food factories are closed. Logistics is severely delayed. Food shortage in the poorest part is getting worse. People are more willing to risk their lives in seeking for food. Violence is starting to take holds in some poorest parts in Africa. Rich countries are focusing on their own pandemic. As a result, the poorest countries are not getting enough help. There is not enough testing or medicine. Also, many volunteering doctors have to be sent back to their home country. Therefore, getting treated for COVID 19 is becoming very difficult. On top of that, here is more bad news. The pandemic is dragging the Global economy into recession. This will inevitably lead to an increase of poverty. World Bank has made an estimation. That is, the pandemic will push forty 9 million people into extreme poverty this year. We will see an increase of poverty across Africa and in populated countries such as India and China. The pandemic will set the poverty rate to increase for the first time since nineteen ninety eight. Nobody could foresee this setback. But it is a challenge for all of us. We must act now. The whole world must work together. A cure for COVID 19 is crucial at the moment. But fighting poverty should be a constant focus.

<u>Difficult Speech</u>

<u>Chunk 1</u>

I learnt something shocking last month when I was watching a YouTube video where
a social science student interviewed random people on the high street about whether they think we are to enter a trust crisis. The interviewer who I think is too young to be cynical started by saying that it seems that there has never been a better moment than now to bring up the topic of mistrust as we are becoming increasingly dubious about so many things that we used to have faith in. According to interviewer's data, one of the key expressions of the past decade is "pseudo kindness". I found the term demeaning. However, my mother who the harshness of life has inured watched the video with me

and eulogised the show as what she believes is that it is paramount to remind people that our time is defined by treachery which is unseen by any other times. I am always a fervent person in believing in goodness. So, I disparaged the idea of reminding people of the opposite. But people do tend to be despondent in difficult times. Even my boss who is the most sanguine person that I know deduced thatwe shall be extra cautious when it comes to trust. I did not bother with ruminating on the topic of trust until recently. Last week, an elderly friend of mine who the cruelty of war has traumatised called me, discussing an issue that I believe is a delicate matter. He said that he was deeply perturbed about his youngest son who has difficulty in making friends because of trust issues. And he would like me who is experienced with teenagers to be didactic with his son about trusting others. The boy is not apathy in others. Instead, he has a zeal for the world. He loves spending time in the garden to botanise the plants. And he is enthused by the topic of geography. Despite his young age, he has got the dexterity for painting. In fact, he is a very genteel young man. His qualities make his unwillingness for friendship even more unfathomable. During our conversation, he confided the reason to me. He was extorted by his best mate. The boy who we think highly of does not hold any rancour towards the misconduct of his mate, which would have been a better situation because instead he has relinquished his pursuit in friendship. He cannot emancipate from the fear of getting hurt like that. His father whom I have told the reason to took the boy to a child specialist who is very famous in helping children regain confidence. What is exhilarating to know is that the boy has made progress and that he begins to hanker for friendship again. The sad story of the boy should pose as a harsh warning that what takes a second to ruin could take much longer to restore, and trust is one of them.

Chunk 2

Like many human instincts, trust is not palpable. But it is by no means a delusion. In fact, trust infiltrates many aspects of our lives. Perhaps one of the best ways to prove the existence of trust is to get swindled. What I am suggesting here is not that we should all get out and become voluntary victims of scams or betrayals so that we can comprehend what trust is. The lesson will come to you regardless of what a vigilant person that you think of yourself. When my generation was growing up, the life that we know today is very different from the old situation where what we perhaps would call tossing stones at a tree entertaining. Today, we live in a culture where the Internet is pretty puissant in manipulating our ideas and behaviours. Without a doubt, the Internet is a grandiose invention. It seemed whimsical at the beginning. But it turns out to be a superlative tool. It is so efficacious in making our lives easy and convenient. It has been transforming our lives in a perceptible way. While it brings excitement to the tips of our fingers, the Internet has also introduced an environment that I would describe as “if you open your window there would be flies in your house”. Whoever uses the Internet would find it familiar that we all aspire to not to be tempted, and yet few attain. Unfortunately, some of the temptations are meant to be malevolent. And they are not fastidious at all about age. My eleven year old son has somehow developed a penchant in learning all sorts of nonsense from the Internet. A few weeks ago, he finally got duped. The boy who leery is what he is considered has lost his prudence when he bought some fraudulent tutorial videos that allege to prepare you to become a tycoon in the stock market. Like what all conmen on the Internet would do to trick you, what they did as the first step to lure my son was to catch his eyes by showing him a free introductive video. The video is absolutely galvanising in many ways. It is resplendent with colours and animations. The speaker who I would imprint as an intelligent economist who happens to have grasped the tricks to abet children spoke with confidence. He was friendly but extremely obsequious. I would totally find it worthy of investing if I watched the video when I was ten. What I find besetting is not just the fact that my son was coaxed which lead to a financial loss, but that he has already experienced one of the darkest sides of human beings. The Internet will always contain unsolicited and malicious contents online. I wonder if it will be a new normal where parents who we think should safeguard the children would need to expose them to the internet to teach them the danger of being credulous.

Appendix E (Glossary for Priming)

Glossary for Easy Speech (Chunk 1)

1. Poorest
2. NGOs
3. Luxury
4. Opposite
5. Privilege
6. Cambodia
7. Poverty
8. Medicine
9. Accomplish
10. Agricultural
11. Primary
12. Nurture
13. Million
14. Crops
15. Witness
16. Volunteer
17. Demanding
18. Fight
19. Hunger
20. Percent
21. Thousand
22. Travel
23. Struggle
24. Transition
25. Vaccination
26. Transform

Glossary for Easy Speech (Chunk 2)

1. Billion
2. Violence
3. Delayed
4. Extreme

5. suffer
6. Tragic
7. shortage
8. Vulnerable
9. Estimation
10. foresee
11. Infection
12. Fragile
13. Theft
14. Challenge
15. Achievement
16. Spread
17. Logistics
18. Inevitably
19. Setback
20. Cure
21. Constant
22. attacked
23. Seeking
24. treated
25. Recession
26. Pandemic

<u>Glossary for Difficult Speech (Chunk 1)</u>

1. **hanker [ˈhæŋkə] 渴望**
   - We all **hanker** for love.
   - 我们都**渴望**爱。

2. **zeal [ziːl] 热情**
   - He always has such **zeal** in playing the piano.
   - 他总是对弹钢琴充满**热情**。

3. **eulogised [ˈjuːlədʒaɪzd] 称赞**
   - The President **eulogised** the bravery of the soldiers.

- 总统**称赞**了士兵的勇敢行为。

4. **treachery [ˈtretʃəri] 背叛**

- He was deeply hurt by his friend’s **treachery**.
- 他由于朋友的**背叛**而深深受伤。

5. **unfathomable [ʌnˈfæðəməbl] 无法理解的**

- The equation is too complicated, it is **unfathomable**.
- 这个公式太复杂了，简直**无法理解**。

6. **genteel [dʒenˈtiːl] 有教养的**

- He has always been a **genteel** person.
- 他一直都是一个**有教养的**人。

7. **demeaning [dɪˈmiːnɪŋ] 有损人格的**

- This is a **demeaning** TV show.
- 这个电视节目**有损人格**。

8. **fervent [ˈfɜːvənt] 充满热诚**

- I am a very **fervent** person in protecting wild animals.
- 我对于保护野生动物**充满热诚**。

9. **dexterity [dekˈsterəti] 本领、技能、能力**

- The little girl is too young to have the **dexterity** to type.
- 这个小女孩年纪太小，还不具备打字的**能力**。

10. **disparaged [dɪˈspærɪdʒd] 批判**

- She **disparaged** how the covid-19 pandemic is handled in the UK.
- 她**批判**了英国抗击新冠大流行病的方式。

11. **apathy [ˈæpəθi] 冷淡、冷漠**

- We can’t show **apathy** to others.
- 我们不能对他人**冷漠**。

12. **despondent [dɪˈspɒndənt] 消极**

- The soldiers feel **despondent** after they were defeated by the enemy.
- 在被敌军打败之后，士兵们感到**消极**。

13. pseudo [sjuːdəʊ] 伪的

- There are many **pseudo-science** stuff, such as detoxing footbath.
- 有许多**伪科学**的事情，比如排毒足浴。

14. ruminating [ˈruːmɪneɪtɪŋ] 思考

- I have been **ruminating** the issue of poverty.
- 我一直在**思考**贫困这个话题。

15. sanguine [ˈsæŋgwɪn] 乐观的

- These young **sanguine** sailors are ready for the drill.
- 这些年轻**乐观的**水手已经准备好加入演习了。

16. confided [kənˈfaɪd] 向…吐露

- I **confided** my secret to my best friend.
- 我向最好的朋友**吐露**了秘密。

17. didactic [daɪˈdæktɪk] 说教的 （教导）

- He likes to be didactic about physics.
- 他喜欢**教导**别人物理。

18. inured [ɪˈnjʊrd] 使…习惯

- Life has **inured** me to anything.
- 生活**使我看惯了**一切。

19. deduced [dɪˈdjuːsd] 认为（本意为“推断”）

- My grandfather **deduced** that I am the one who bought him the cake.
- 我外公**认为**蛋糕是我买的。

20. perturbed [pəˈtɜːbd] 困扰

- Linda is deeply **perturbed** by the criticism from her boss.
- Linda 因为收到了老板的批评而深感**困扰**。

21. botanise [ˈbɒtənaɪz] 观察植物

- Last week, we went to the field to **botanise** plants.
- 上周，我们去了野外**观察植物**。

22. enthused [ɪnˈθjuːzd] 使…充满热情

- The movie has **enthused** him in exploring the space.
- 这个电影让他对探索太空**充满热情**。

23. rancour [ˈræŋkə] 怨恨

- He showed no **rancour** to the perpetrators.
- 他对肇事者没有心怀**怨恨**。

24. relinquished [rɪˈlɪŋkwɪʃd] 放弃

- He has **relinquished** his passion in music.
- 他**放弃了**对音乐的热情。

25. emancipate [ɪˈmænsɪpeɪt] 从…中解放

- We must **emancipate** from the sorrow in the past.
- 我们要**从**过去的伤痛**中解放**出来。

26. extorted [ɪkˈstɔːt] 被勒索

- He was **extorted** last year.
- 他去年**被勒索**了。

<u>**Glossary for Difficult Speech (Chunk 2)**</u>

1. Attain [əˈteɪn] 做到

- All athletes hope to be Olympic champions, but few **attain**.
- 所有运动员都渴望成为奥运冠军，但是能**做到**的人很少。

2. Prudence [ˈpruːdns]谨慎

- You should always have **prudence** with spending money.
- 在花钱的时候，你应该保持谨慎。

3. Duped [dju:pt] 被骗了

- We were **duped** last month.
- 我们上个月**被骗了**。

4. Credulous [kredju:ləs]容易相信别人/轻信

- He is always **credulous**.
- 他总是很**容易相信别人/轻信**。

5. Fastidious [fəsti:diəs] 挑剔的

- Japanese people are **fastidious** about details.
- 日本人对细节是很**挑剔的**。

6. Abet [əbet] 怂恿

- That man tried to abet young people to **steal**.
- 那个人**怂恿**年轻人盗窃。

7. Obsequious [əbˈsi:kwiəs]谄媚的

- He is always **obsequious** around powerful people.
- 他在有权有势的人面前总是一副**谄媚的**样子。

8. Besetting [biˈsetiŋ] 令人不安的

- That Edward is sent to the hospital is a **besetting** news.
- 爱德华被送进医院的消息**令人不安**。

9. Puissant [ˈpju: isant] 非常善长的、强有力的、强大的

- Supercomputer is **puissant** in cloud computing.
- 超级计算机在云计算方面是很**强大的**。

10. Grandiose[ˈgrændiəʊs] 伟大的

- Mobile phones are **grandiose** inventions.
- 手机是**伟大的**发明。

11. Whimsical [ˈwɪmzɪkl]奇怪的

- This video game seems **whimsical**.
- 这个电游有点**奇怪**。

12. Palpable [ˈpælpəbl 可触摸到的

- The excitement at Disneyland is almost **palpable**.
- 我们甚至**可以触摸到**人们在迪士尼乐园里的兴奋。

13. Superlative [suːˈpɜːlətɪv]最佳的

- She is a **superlative** actress.
- 她是**最佳**女演员。

14. Allege [əˈledʒ] 声称

- This book **alleges** to help you understand love.
- 这本书**声称**可以帮你读懂爱。

15. Leery [ˈlɪəri] 十分小心/多疑的

- He is such a **leery** person that he never trusts anyone.
- 他是一个**十分小心/多疑**的人，不相信任何人。

16. Penchant [ˈpõʃõ] 热爱/强烈的兴趣*

- He has a **penchant** for Champagne.
- 他对香槟有**强烈的兴趣**。

17. Galvanising [ˈgælvənaɪzɪŋ] 吸引眼球的、吸睛的

- The poster of that super model is absolutely galvanising!
- 那个超模的海报太吸引眼球了。

18. Resplendent [rɪˈsplendənt] 华丽的

- This is a **resplendent** dress.
- 这件裙子很**华丽**。

19. Infiltrates [ˈɪnfɪltreɪts] 渗透

- The special agent seeks to **infiltrate** the gang.
- 一名特工试图**渗透**到黑帮内部。

20. Swindled [ˈswɪndld] 被骗

- My boss was **swindled** last week.
- 我的老板上周**被骗**了。

21. Malevolent [məˈlevələnt] 恶意的

- I received a **malevolent** email.
- 我收到了一封**恶意**邮件。

22. Efficacious [ˌefɪˈkeɪʃəs] 有效的

- A change in diet would be as **efficacious** as taking the medicine.
- 改善饮食和用药是一样**有效的**。

23. Perceptible [pəˈseptəbl] 可察觉的、可感受到的

- There has been a perceptible decline in public confidence towards the government.
- 我们**可以感受到**大众对政府的信任正在下降。

24. Coaxed [kəʊksd]被哄骗的

- He was **coaxed** into the boring work.
- 他**被骗**来做这个无聊的工作。

25. Unsolicited [ˌʌnsəˈlɪsɪtɪd] 不恰当的

- I receive **unsolicited** junk mails every day.
- 我每天都会收到**内容不恰当的**邮件。

26. Delusion [dɪˈluːʒn]错觉

- That the summer feels colder than the winter is a **delusion**.
- 夏天比冬天冷是**错觉**。

Appendix F (Marking Sheet)

Easy Speech

Chunk 1

Sentence segments	GA?	score	Sentence Number
1. A few 2. years 3. ago 4. developed 5. countries 6. made a donation.			1
1. They 2. gave 3. money to 4. the poorest 5. countries .			2
1. The goal 2. Was 3. to solve 4. one of 5. the world's 6. biggest 7. problems.			3
1. That 2. Is 3. hunger .			4
1. Many 2. NGOs 3. volunteered,			5

4. too.			
1. They 2. helped 3. the poor 4. countries 5. to receive 6. and 7. manage 8. the funds.			6
1. I 2. work 3. for 4. one of 5. the NGOs.			7
1. The job 2. is 3. very 4. physically 5. demanding.			8
1. I 2. live 3. in 4. the UK.			9
1. But 2. my 3. clients 4. are 5. mainly 6. in 7. Africa 8. and 9. Asia.			10

1. I 2. have to 3. adapt to 4. their 5. working 6. hours.			11
1. Regular 2. sleeping 3. hours 4. to 5. me 6. is 7. a luxury.			12
1. But 2. there is 3. a privilege 4. from 5. my 6. job.			13
1. That is, 2. I 3. travel 4. a lot.			14
1. My 2. job 3. does not 4. usually 5. take 6. me 7. to 8. rich 9. countries.			15
1. It			16

2. sends 3. me 4. to 5. the opposite.			
1. To understand 2. poverty, 3. you 4. have to 5. be 6. there.			17
1. I 2. have been to 3. many 4. poor 5. countries.			18
1. I 2. witnessed 3. the daily 4. struggles 5. for 6. food 7. and 8. medicine.			19
1. I 2. saw 3. people 4. fight 5. to death 6. for 7. clean 8. water.			20
1. Take 2. Cambodia			21

3. for 4. example.			
1. I 2. lived 3. in 4. Cambodia 5. between 6. two thousand and six 7. and 8. two thousand and nine.			22
1. Cambodia 2. has been 3. one of 4. the poorest 5. countries.			23
1. According to 2. World Bank, 3. Cambodia 4. had 5. a poverty rate of 6. forty seven percent 7. in 8. two thousand and seven.			24
1. That 2. is 3. a really 4. big 5. number.			25
1. I 2. have 3. more			26

4. figures 5. for 6. you.			
1. Ten years 2. ago, 3. forty percent 4. of 5. the population 6. did not 7. have 8. access to 9. life saving 10. medicines.			27
1. The average 2. life expectancy 3. was 4. sixty 5. for 6. men 7. and 8. sixty two 9. for 10. women;			28
1. Over 2. five percent 3. of 4. women 5. died 6. during 7. labour;			29
1. Nearly 2. nine percent 3. of 4. Cambodian			30

5. children 6. died 7. before 8. age of five;			
1. Most 2. people 3. could 4. only 5. have 6. one 7. meal 8. each 9. day;			31
1. Clean 2. water 3. was 4. a dream 5. for 6. most 7. of 8. them;			32
1. Primary 2. education 3. was 4. next to 5. none.			33
1. I 2. only 3. lived 4. in 5. Cambodia for 6. three 7. years.			34

1. It 2. was 3. a short 4. period of time.			35
1. But 2. I 3. witnessed 4. some 5. significant 6. transition.			36
1. With the donation 2. and 3. other forms 4. of 5. help, 6. middle 7. Cambodia 8. accomplished 9. great 10. achievements.			37
1. Today, 2. most 3. people 4. in 5. middle 6. Cambodia 7. still 8. live 9. in 10. rural 11. areas.			38
1. But 2. eighteen percent 3. of			39

4. the local 5. population 6. managed 7. to escape 8. poverty			
1. Modern 2. agricultural 3. machines 4. were donated to 5. the area.			40
1. A large number of 2. local 3. farmers 4. learnt to 5. use 6. them.			41
1. Better 2. crops 3. also 4. were brought 5. to the area.			42
1. So 2. they 3. produced 4. much more 5. food.			43
1. Almost 2. all 3. adults 4. can 5. have 6. two 7. meals 8. a day.			44

1. Children 2. can 3. even 4. have 5. three 6. meals.			45
1. NGOs 2. helped 3. to build 4. hospitals.			46
1. Most 2. women 3. can 4. give birth 5. in 6. a clean 7. and 8. safe 9. environment.			47
1. Doctors 2. and 3. health 4. workers 5. are helping 6. mothers 7. to nurture 8. their 9. babies.			48
1. Free 2. Vaccinations 3. for 4. life threatening 5. and			49

6. infectious 7. diseases 8. are becoming 9. available.			
1. Local 2. authorities 3. also 4. are working 5. with 6. NGOs 7. to 8. promote 9. education			50
1. Primary 2. schools 3. and 4. vocational 5. centres 6. have been set up 7. across 8. the area.			51
1. All 2. courses 3. are 4. free 5. and 6. taught 7. by 8. international 9. volunteers.			52
1. These 2. improvements 3. transformed 4. the area.			53

1. Millions of 2. people 3. are living 4. a better 5. life.			54
1. There is 2. still 3. a long 4. way 5. to go.			55
1. But 2. people 3. now 4. have 5. hopes.			56
1. And we 2. all 3. are trying 4. hard 5. to 6. give 7. them 8. more 9. hopes.			57

Chunk 2

Sentence segments	GA?	Score	
1. In 2. nineteen ninety, 3. thirty six percent 4. of			1

5. the world's population 6. lived in 7. poverty.			
1. That 2. was 3. nearly 4. two billion 5. people.			2
1. Most of 2. them 3. lived 4. under 5. extreme 6. poverty 7. at that time.			3
1. In other words, 2. they 3. lived 4. under 5. less than fifty 6. pence 7. a 8. day.			4
1. A bottle of 2. coca cola 3. would cost 4. you 5. thirty five 6. pence 7. in 8. the UK 9. back then.			5

1. The world 2. has been fighting 3. poverty 4. in many ways.			6
1. Owing to 2. the hard 3. work, 4. the poverty 5. rate 6. dropped 7. to 8. ten percent 9. today.			7
1. This 2. is 3. a 4. great 5. achievement.			8
1. The world 2. is 3. indeed 4. in 5. a better 6. place.			9
1. But 2. ten percent 3. of 4. the population 5. is 6. still 7. a 8. huge 9. number.			10

1. Over 2. seven hundred million 3. people 4. are living on 5. less than 6. one 7. pound 8. a 9. day.			11
1. That 2. is 3. beyond 4. tragic.			12
1. Poverty 2. does not 3. just 4. mean 5. no 6. money.			13
1. It 2. hurts 3. human 4. race 5. several 6. ways.			14
1. Needless to say, 2. poverty 3. damage 4. our 5. health.			15
1. In 2. the poorest			16

3. areas, 4. people 5. suffer from 6. bad 7. living 8. conditions.			
1. Most of 2. them 3. don’t have 4. windows 5. on 6. the wall.			17
1. The shortage 2. of 3. fresh 4. air 5. can 6. easily 7. cause 8. lung 9. infections.			18
1. Toilets 2. are 3. also 4. a 5. luxury.			19
1. The lack of 2. proper 3. toilets 4. leads to 5. the spread 6. of 7. life-threatening 8. illness.			20

1. The fragile 2. houses 3. easily 4. are ruined 5. by 6. rain, 7. storms, 8. and earthquakes.			21
1. Some of 2. the poorest 3. have to 4. live 5. in 6. a tent 7. or 8. under 9. a shelter.			22
1. Anything 2. they 3. own 4. is 5. at the risk 6. of 7. theft.			23
1. They 2. could 3. be attacked 4. or 5. even 6. killed 7. by 8. wild 9. animals.			24

1. Children 2. under 3. poverty 4. sometimes 5. do not 6. get 7. education.			25
1. They 2. very 3. rarely 4. realise 5. their 6. potential.			26
1. The poorest 2. population 3. are 4. more vulnerable than 5. the rest 6. of 7. us.			27
1. We 2. are 3. in 4. the COVID nineteen pandemic 5. now.			28
1. It 2. is 3. hitting 4. the poorest 5. people 6. particularly 7. hard.			29

1. Many 2. food 3. factories 4. are closed.			30
1. Logistics 2. Is 3. severely 4. delayed.			31
1. Food 2. shortage 3. in 4. the poorest 5. part 6. is getting 7. worse.			32
1. People 2. are 3. more willing 4. to risk 5. their 6. lives 7. in 8. seeking 9. for 10. food.			33
1. Violence 2. is starting 3. to take holds 4. in 5. some 6. poorest 7. parts 8. in			34

9. Africa.			
1. Rich 2. countries 3. are focusing 4. on 5. their 6. own 7. pandemic.			35
1. As a result, 2. the poorest 3. countries 4. are not 5. getting 6. enough 7. help.			36
1. There is 2. not 3. enough 4. testing 5. or 6. medicine.			37
1. Also, 2. many 3. volunteering 4. doctors 5. have to 6. be sent 7. back 8. to 9. their 10. home 11. country.			38
1. Therefore,			39

2. getting treated 3. for 4. COVID nineteen 5. is becoming 6. very 7. difficult.			
1. On top of that, 2. here 3. is 4. more 5. bad 6. news.			40
1. the pandemic 2. is dragging 3. the Global 4. economy 5. into 6. recession			41
1. This 2. inevitably 3. will lead to 4. an increase 5. of 6. poverty.			42
1. World Bank 2. has made 3. an 4. estimation.			43
1. That is, 2. the pandemic 3. will push 4. forty nine million 5. people			44

6. into 7. extreme 8. poverty 9. this 10. year.			
1. We 2. will see 3. an 4. increase 5. of 6. poverty 7. across 8. Africa 9. and 10. in 11. populated 12. countries 13. such as 14. India 15. and 16. China.			45
1. The pandemic 2. will set 3. the poverty 4. rate 5. to increase 6. for the first time 7. since 8. nineteen ninety eight.			46
1. Nobody 2. could 3. foresee 4. this 5. setback.			47

1. But 2. it 3. is 4. a 5. challenge 6. for 7. all 8. of 9. us.			48
1. We 2. must 3. act 4. now.			49
1. The whole 2. world 3. must 4. work 5. together.			50
1. A 2. cure 3. for 4. COVID nineteen 5. is 6. crucial 7. at the moment.			51
1. But 2. fighting 3. poverty 4. should be 5. a 6. constant 7. focus.			52

Difficult Speech

Chunk 1

Sentence segments	GA?	score	Sentence number
1. I 2. learnt 3. something 4. shocking 5. last 6. month when 7. I 8. was watching 9. a YouTube 10. video where 11. a social 12. science 13. student 14. interviewed 15. random 16. people 17. on 18. the high street 19. about 20. whether 21. they 22. think 23. we 24. are 25. to enter 26. a trust 27. crisis.			1
1. The interviewer who 2. I			2

3. Think 4. Is 5. too 6. young 7. to be 8. cynical 9. started 10. by 11. saying that 12. it seems that 13. there 14. has never been 15. better than now 16. a moment 17. to bring up 18. the topic of 19. mistrust 20. as 21. we 22. are becoming 23. increasingly 24. dubious 25. about 26. so many 27. things that 28. we 29. used 30. to have 31. faith 32. in.			
1. According to 2. interviewer's 3. data, 4. one of 5. the key 6. expressions 7. of	左		3

8. the past 9. decade 10. is 11. “pseudo 12. kindness”.			
1. I 2. Found 3. the term 4. demeaning.			4
1. However, 2. my 3. mother who 4. the harshness 5. of 6. life 7. has inured 8. watched 9. the video 10. with 11. me 12. and 13. eulogised 14. the show 15. as what 16. she 17. believes 18. is that 19. it is paramount 20. to remind 21. people that 22. our 23. time 24. is defined 25. by 26. treachery which 27. is unseen			5

28. by 29. any other 30. times.			
1. I 2. am 3. always 4. a fervent 5. person 6. in believing in 7. goodness.			6
1. So, 2. I 3. disparaged 4. the idea of 5. reminding 6. people of 7. the opposite.			7
1. But 2. people 3. do tend to be 4. despondent 5. in 6. difficult 7. times.			8
1. Even 2. my 3. boss who 4. is 5. the most sanguine 6. person that 7. I 8. know 9. deduced that 10. we			9

11. shall be 12. extra 13. cautious when 14. it 15. comes 16. to trust.			
1. I 2. did not 3. bother with 4. ruminating 5. on 6. the topic of 7. trust 8. until 9. recently.			10
1. Last 2. week, 3. an elderly 4. friend 5. of 6. mine who 7. the cruelty 8. of 9. war 10. has traumatised 11. called 12. me, 13. discussing 14. an issue that 15. I 16. believe 17. is 18. a delicate 19. matter.			11
1. He			12

2. said that 3. he 4. was 5. deeply 6. perturbed 7. about 8. his 9. youngest 10. son who 11. has 12. difficulty 13. in 14. making 15. friends 16. because of 17. trust 18. issues.			
1. And he 2. would like 3. me who 4. is experienced 5. with 6. teenagers 7. to be didactic 8. with 9. his 10. son 11. about 12. trusting 13. others.			13
1. The 2. boy 3. is not 4. apathy 5. in 6. others.			14
1. Instead,			15

2. he 3. has 4. a zeal 5. for 6. the world			
1. He 2. loves 3. spending 4. time 5. in 6. the garden 7. to botanise 8. the plants.			16
1. And he 2. is enthused 3. by 4. the topic of 5. geography.			17
1. Despite 2. his 3. young 4. age, 5. he 6. has got 7. the dexterity 8. for painting.			18
1. In fact, 2. he 3. is 4. a very 5. genteel 6. young 7. man.			19

1. His 2. qualities 3. make 4. his 5. unwillingness 6. for 7. friendship 8. even more unfathomable.			20
1. During 2. our 3. conversation, 4. he 5. confided 6. the reason 7. to me.			21
1. He 2. was extorted 3. by 4. his 5. best 6. mate.			22
1. The boy who 2. we 3. think of 4. highly 5. does not 6. hold 7. any 8. rancour 9. towards 10. the misconduct 11. of 12. his 13. mate, 14. which would have been			23

15. a better 16. situation 17. because 18. instead 19. he 20. has relinquished 21. his 22. pursuit 23. in 24. friendship.			
1. He 2. cannot 3. emancipate 4. from 5. the fear 6. of 7. getting 8. hurt 9. like 10. that.			24
1. His 2. father whom 3. I 4. have told 5. the reason to 6. took 7. the boy 8. to 9. a child 10. specialist 11. who 12. is 13. very 14. famous 15. in 16. helping			25

17. children 18. regain 19. confidence.			
1. What is exhilarating 2. to know 3. is that 4. the boy 5. has made 6. progress 7. and that 8. he 9. begins 10. to hanker 11. for 12. friendship 13. again.			26
1. The sad 2. story 3. of 4. the boy 5. should 6. pose 7. as 8. a harsh 9. warning that 10. what takes a 11. second 12. to ruin 13. could 14. take 15. much longer 16. to restore, 17. and trust 18. is 19. one of 20. them.			27

Chunk 2			
sentence	GA?	score	
1. Like 2. many 3. human 4. instincts, 5. trust 6. is not 7. palpable.			1
1. But 2. it 3. is 4. by no means 5. a delusion.			2
1. In fact, 2. trust 3. infiltrates 4. many 5. aspects 6. of 7. our 8. lives.			3
1. Perhaps 2. one of 3. the best 4. ways 5. to prove 6. the existence 7. of 8. trust 9. is to 10. get			4

11. swindled.			
1. What I 2. am suggesting 3. here 4. is not that 5. we 6. should 7. all 8. get out 9. and 10. become 11. voluntary 12. victims 13. of 14. scams 15. or 16. betrayals 17. so that 18. we 19. can 20. comprehend 21. what 22. trust 23. is.			5
1. The lesson 2. will come 3. to you 4. regardless of 5. what 6. a vigilant 7. person that 8. you 9. think of 10. yourself.			6
1. When 2. my			7

3. generation 4. was growing up, 5. the life that 6. we 7. know 8. today 9. is 10. very 11. different from 12. the old 13. situation where 14. we 15. perhaps 16. would call 17. tossing at 18. stones 19. a tree 20. entertaining.			
1. Today, 2. we 3. live 4. in 5. a culture where 6. the Internet 7. is 8. pretty 9. puissant 10. in 11. manipulating 12. our 13. ideas 14. and 15. behaviours.			8
1. Without a doubt, 2. the Internet 3. is			9

4. a 5. grandiose 6. invention.			
1. It 2. seemed 3. whimsical 4. at the beginning.			10
1. But 2. it 3. turns out 4. to be 5. a 6. superlative 7. tool.			12
1. It 2. is 3. so efficacious 4. in 5. making 6. our 7. lives 8. easy 9. and 10. convenient.			12
1. It 2. has been transforming 3. our 4. lives 5. in 6. a perceptible 7. way.			13
1. While 2. it 3. brings			14

4. excitement 5. to the tips 6. of 7. our 8. fingers, 9. the Internet 10. also 11. has introduced 12. an 13. environment that 14. I 15. would describe 16. as 17. “if 18. you 19. open 20. your 21. window 22. there would be 23. flies 24. in 25. your 26. house”.			
1. Whoever 2. uses 3. the Internet 4. would find 5. it 6. familiar that 7. we 8. all 9. aspire to 10. not to 11. be tempted, 12. and yet 13. few 14. attain.			15

1. Unfortunately, 2. some of 3. the temptations 4. are meant 5. to be 6. malevolent.			16
1. And they 2. are not 3. fastidious 4. at all 5. about 6. age.			17
1. My 2. eleven year old 3. son 4. somehow 5. has developed 6. a penchant 7. in 8. learning 9. all sorts of 10. nonsense 11. from 12. the Internet.			18
1. A few 2. weeks 3. ago, 4. he 5. finally 6. got duped.			19
1. The boy who 2. Leery 3. is what 4. he			20

5. is considered 6. has lost 7. his 8. prudence 9. when 10. he 11. bought 12. some 13. fraudulent 14. tutorial 15. videos that 16. allege to 17. prepare 18. you 19. to become 20. a 21. tycoon 22. in 23. the stock 24. market.			
1. Like 2. what all 3. conmen 4. on 5. the Internet 6. would do 7. to 8. trick 9. you, 10. what they 11. did 12. as 13. the first 14. step 15. to lure 16. my 17. son			21

18. was 19. to catch 20. his 21. eyes 22. by 23. showing 24. him 25. a 26. free 27. introductive 28. video.			
1. The video 2. Is 3. Absolutely 4. Galvanising 5. in 6. many 7. ways.			22
1. It 2. is 3. resplendent 4. with 5. colours 6. and 7. animations.			23
1. The speaker who 2. I 3. would imprint 4. as 5. an 6. intelligent 7. economist who 8. happens 9. to have grasped 10. the tricks			24

11. to abet 12. children 13. spoke 14. with 15. confidence.			
1. He 2. was 3. friendly 4. but 5. extremely 6. obsequious.			25
1. I 2. totally 3. would find 4. it 5. worthy of 6. investing 7. if 8. I 9. watched 10. the video 11. when 12. I 13. was 14. ten.			26
1. What I 2. find 3. besetting 4. is not 5. just 6. the fact that 7. my 8. son 9. was coaxed which 10. lead to			27

11. a 12. financial 13. loss, 14. but that 15. he 16. already 17. has experienced 18. one of 19. the darkest 20. sides 21. of 22. human beings.			
1. The Internet 2. will contain 3. always 4. unsolicited 5. and 6. malicious 7. contents 8. online.			28
1. I 2. Wonder 3. if 4. it 5. will be 6. a 7. new 8. normal where 9. parents who 10. we 11. think 12. should 13. safeguard 14. the 15. children 16. would need to			29

17. expose 18. them 19. to 20. the internet 21. to 22. teach 23. them 24. the danger 25. of 26. being credulous.			

Chunk 2

Sentence segments	GA?	score	Sentence number
8. Like 9. many 10. human 11. instincts, 12. trust 13. is not 14. palpable.			1
6. But 7. it 8. is 9. by no means 10. a delusion.			2
9. In fact, 10. trust 11. infiltrates 12. many 13. aspects 14. of 15. our 16. lives.			3

12. Perhaps 13. one of 14. the best 15. ways 16. to prove 17. the existence 18. of 19. trust 20. is to 21. get 22. swindled.			4
24. What I 25. am suggesting 26. here 27. is not that 28. we 29. should 30. all 31. get out 32. and 33. become 34. voluntary 35. victims 36. of 37. scams 38. or 39. betrayals 40. so that 41. we 42. can 43. comprehend 44. what 45. trust 46. is.			5
11. The lesson			6

12. will come 13. to you 14. regardless of 15. what 16. a vigilant 17. person that 18. you 19. think of 20. yourself.			
21. When 22. my 23. generation 24. was growing up, 25. the life that 26. we 27. know 28. today 29. is 30. very 31. different from 32. the old 33. situation where 34. we 35. perhaps 36. would call 37. tossing at 38. stones 39. a tree 40. entertaining.			7
16. Today, 17. we 18. live 19. in 20. a culture where 21. the Internet 22. is 23. pretty			8

24. puissant 25. in 26. manipulating 27. our 28. ideas 29. and 30. behaviours.			
7. Without a doubt, 8. the Internet 9. is 10. a 11. grandiose 12. invention.			9
5. It 6. seemed 7. whimsical 8. at the beginning.			10
8. But 9. it 10. turns out 11. to be 12. a 13. superlative 14. tool.			12
11. It 12. is 13. so efficacious 14. in 15. making 16. our 17. lives 18. easy 19. and 20. convenient.			12

8. It 9. has been transforming 10. our 11. lives 12. in 13. a perceptible 14. way.			13
27. While 28. it 29. brings 30. excitement 31. to the tips 32. of 33. our 34. fingers, 35. the Internet 36. also 37. has introduced 38. an 39. environment that 40. I 41. would describe 42. as 43. “if 44. you 45. open 46. your 47. window 48. there would be 49. flies 50. in 51. your 52. house”.			14
15. Whoever 16. uses 17. the Internet			15

18. would find 19. it 20. familiar that 21. we 22. all 23. aspire to 24. not to 25. be tempted, 26. and yet 27. few 28. attain.			
7. Unfortunately, 8. some of 9. the temptations 10. are meant 11. to be 12. malevolent.			16
7. And they 8. are not 9. fastidious 10. at all 11. about 12. age.			17
13. My 14. eleven year old 15. son 16. somehow 17. has developed 18. a penchant 19. in 20. learning 21. all sorts of 22. nonsense 23. from 24. the Internet.			18

7. A few 8. weeks 9. ago, 10. he 11. finally 12. got duped.			19
25. The boy who 26. Leery 27. is what 28. he 29. is considered 30. has lost 31. his 32. prudence 33. when 34. he 35. bought 36. some 37. fraudulent 38. tutorial 39. videos that 40. allege to 41. prepare 42. you 43. to become 44. a 45. tycoon 46. in 47. the stock 48. market.			20
29. Like 30. what all 31. conmen 32. on 33. the Internet 34. would do	左		21

35. to 36. trick 37. you, 38. what they 39. did 40. as 41. the first 42. step 43. to lure 44. my 45. son 46. was 47. to catch 48. his 49. eyes 50. by 51. showing 52. him 53. a 54. free 55. introductive 56. video.			
8. The video 9. Is 10. Absolutely 11. Galvanising 12. in 13. many 14. ways.			22
8. It 9. is 10. resplendent 11. with 12. colours 13. and 14. animations.			23

16. The speaker who 17. I 18. would imprint 19. as 20. an 21. intelligent 22. economist who 23. happens 24. to have grasped 25. the tricks 26. to abet 27. children 28. spoke 29. with 30. confidence.			24
7. He 8. was 9. friendly 10. but 11. extremely 12. obsequious.			25
15. I 16. totally 17. would find 18. it 19. worthy of 20. investing 21. if 22. I 23. watched 24. the video 25. when 26. I 27. was 28. ten.			26

23. What I 24. find 25. besetting 26. is not 27. just 28. the fact that 29. my 30. son 31. was coaxed which 32. lead to 33. a 34. financial 35. loss, 36. but that 37. he 38. already 39. has experienced 40. one of 41. the darkest 42. sides 43. of 44. human beings.			27
9. The Internet 10. will contain 11. always 12. unsolicited 13. and 14. malicious 15. contents 16. online.			28
27. I 28. Wonder 29. if 30. it 31. will be			29

32. a 33. new 34. normal where 35. parents who 36. we 37. think 38. should 39. safeguard 40. the 41. children 42. would need to 43. expose 44. them 45. to 46. the internet 47. to 48. teach 49. them 50. the danger 51. of 52. being credulous.			

Appendix G

GA Measurements for Cohorts 2013 - 2016

2013

Interpreter	***Total GA counts***	***GA counts/100 words***	***GA counts/100 seconds***	***Total GA Length***	***GA Proportion (%)***
1	4	0.76	1.80	4.8	2.16
2	7	1.33	3.15	4.59	2.07
3	3	0.57	1.35	0.91	0.41
4	6	1.14	2.70	6.35	2.86
5	14	2.65	6.31	10.03	4.52
6	10	1.89	4.50	7.02	3.16
7	13	2.46	5.86	6.44	2.90
8	5	0.95	2.25	3.55	1.60
9	7	1.33	3.15	9.65	4.35
Median	7	1.33	3.15	6.35	2.86
Mean	7.67	1.45	3.45	5.93	2.67
SD	3.87	0.73	1.74	2.88	1.30

2014

Interpreter	***Total GA Counts***	***GA Counts/100 Words***	***GA Counts/100 Seconds***	***Total GA Length***	***GA Proportion (%)***
1	9	1.97	4.13	13.55	6.22
2	11	2.40	5.05	13.83	6.34
3	5	1.09	2.29	3.63	1.67
4	21	4.59	9.63	35.36	16.22
5	10	2.18	4.59	6.73	3.09
Median	10	2.18	4.59	13.55	6.22
Mean	11.2	2.45	5.14	14.62	6.71
SD	5.93	1.30	2.72	12.40	0.06

2015

Interpreter	*Total GA counts*	*GA counts/100 words*	*GA counts/100 seconds*	*Total GA Length*	*GA Proportion (%)*
1	6	1.20	3.09	8.68	4.47
2	18	3.59	9.28	33.05	17.04
3	8	1.59	4.12	4.78	2.46
4	20	3.98	10.31	19.46	10.03
5	4	0.80	2.06	4.01	2.07
6	45	8.96	23.20	81.96	42.25
7	6	1.20	3.09	6.77	3.49
Median	8	1.59	4.12	8.68	4.47
Mean	15.29	3.04	7.88	22.67	11.69
SD	14.52	2.89	7.49	28.14	0.15

2016

Interpreter	***Total GA counts***	***GA counts/100 words***	***GA counts/100 seconds***	***Total GA Length***	***GA Proportion (%)***
1	5	1.11	2.16	7.45	3.21
2	17	3.76	7.33	22.83	9.84
3	24	5.31	10.34	28.86	12.44
4	38	8.41	16.38	49.08	21.26
5	9	1.99	3.88	8.82	3.80
6	25	5.53	10.78	18.2	7.84
7	9	1.99	3.88	3.95	1.70
8	21	4.65	9.05	18.61	8.02
9	40	8.85	17.24	39.1	16.85
Median	21	4.65	9.05	18.61	8.02
Mean	20.89	4.62	9.00	21.88	9.44
SD	12.42	2.75	5.36	15.04	6.51

Appendix H

Master Sheet for GA Data from the Experimental Study (Easy and Difficult Speech)

Master Sheet for GA under the Easy Speech from All Interpreters

Interpreter	***Total GA Counts***	***Total GA Length (Seconds)***	***GA/100 Source Speech Words***	***GA/100 Speech (Seconds)***	***GA Proportion (%)***
1	0	0	0	0	0
2	1	1.44	0.1	0.24	0.34
3	14	17.01	1.43	3.33	4.05
4	0	0	0	0	0
5	3	1.67	0.31	0.71	0.4
6	0	0	0	0	0
7	2	0.63	0.2	0.48	0.15
8	16	12.48	1.64	3.81	2.97
9	0	0	0	0	0
10	0	0	0	0	0
11	2	0.98	0.2	0.48	0.23
12	0	0	0	0	0
13	0	0	0	0	0
14	16	9.9	1.64	3.81	2.36
15	0	0	0	0	0
16	18	10.44	1.84	4.29	2.49
Total	72	54.54	7.36	17.14	NA
Median	0.5	0.32	0.05	0.12	0.08
Mean	4.5	0.76	0.46	1.07	0.81
SD	6.96	5.61	0.71	1.66	1.34

Master Sheet for GA under the Difficult Speech from All Interpreters

Interpreter	***Total GA Counts***	***Total GA Length (Seconds)***	***GA/100 Source Speech Words***	***GA/100 Speech (Seconds)***	***GA Proportion (%)***
1	0	0	0	0	0
2	10	9.08	1.02	2.38	2.16
3	34	57.76	3.07	7.14	13.75
4	4	1.8	0.41	0.95	0.43
5	16	15.09	1.64	3.81	3.59
6	0	0	0	0	0
7	10	7.79	1.02	2.38	1.86
8	18	30.87	1.84	4.29	7.35
9	0	0	0	0	0
10	0	0	0	0	0
11	9	9.68	0.92	2.14	2.31
12	4	3.76	0.41	0.95	0.89
13	0	0	0	0	0
14	59	44.36	6.03	14.05	10.56
15	0	0	0	0	0
16	29	25.4	2.97	6.9	6.05
Total	189	205.59	19.33	45	NA
Median	6.50	5.78	0.67	1.55	1.38
Mean	11.81	1.09	1.21	2.81	3.06
SD	16.09	17.78	1.65	3.83	4.23

Appendix I

Counts and Average Length of GA under Each Sentence Type

Interpreter	***GA Counts (CS)***	***GA Counts (LFW)***	***GA Counts (Both)***	***Avg. GA Length (CS)***	***Avg. GA Length (LFW)***	***Avg. GA Length (Both)***
1	0	0	0	0	0	0
2	4	3	3	1.01	1.15	1.8
3	12	10	12	1.52	1.67	2.32
4	0	3	1	0	0.22	1.14
5	3	7	6	0.89	0.71	1.24
6	0	0	0	0	0	0
7	1	3	6	0.72	0.74	0.81
8	7	4	7	1.28	2.02	1.98
9	5	0	4	1.33	0	3.51
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	2	1	1	0.94	1.12	0
14	0	0	0	0	0	0
15	14	24	21	0.75	0.71	0.79
16	9	8	12	0.74	0.84	1.01
Total	57	63	73	9.18	9.18	14.6
Median	1.5	2	2	0.73	0.47	0.80
Mean	3.56	3.94	4.56	0.57	0.57	0.91
SD	4.63	6.24	6.02	0.56	0.66	4.99

Appendix J

GA Measurements under Longer and Shorter Exposure

Interpreter	*GA Counts*		*Total GA Length*		*GA/100 Words*		*GA/100 Seconds*		*GA Proportion (%)*	
	Shorter	*Longer*	*Shorter*	*Longer*	*Shorter*	*Longer*	*Shorter*	*Longer*	*Shorter*	*Longer*
1	0	0	0	0	0	0	0	0	0	0
2	4	6	2.8	6.3	0.82	0.41	1.9	2.86	1.33	3
3	14	20	28.28	34.4	2.86	4.09	6.67	9.52	13.48	16.44
4	3	1	1.68	0.12	0.61	0.82	1.43	0.48	0.8	0.06
5	0	0	0	0	0	1.02	0	0	0	0
6	0	0	0	0	0	1.23	0	0	0	0
7	4	5	3.4	6.3	0.82	1.43	1.9	2.38	1.62	2.99
8	3	1	2.64	1.12	0.61	1.64	1.43	0.48	1.26	0.53
9	4	12	5.8	9.36	0.82	2.45	1.9	5.71	2.75	4.43
10	0	0	0	0	0	0	0	0	0	0
11	5	13	5.65	25.22	1.02	2.66	2.38	6.19	2.69	12.01
12	4	6	3.68	4.14	0.82	1.23	1.9	2.86	1.75	1.96
13	30	29	21.9	22.33	5.32	5.93	12.38	13.81	9.42	10.63
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	15	14	15	10.36	3.07	2.25	7.14	5.24	7.17	3.68
Total	86	107	90.83	119.65	16.77	25.16	39.03	49.53	NA	NA
Mean	5.38	6.69	5.68	7.48	1.05	1.57	2.44	3.10	2.64	3.48
SD	7.77	8.42	8.28	10.36	1.43	1.58	3.33	3.94	3.83	4.93

Appendix K

LFWs Hit-Rates (%) under Longer and Shorter Priming

Interpreter	***Longer***	***Shorter***
1	69.23	65.38
2	38.46	23.08
3	30.77	30.77
4	38.46	30.77
5	19.23	15.38
6	61.54	61.54
7	46.15	19.23
8	46.15	26.92
9	26.92	61.54
10	42.31	34.62
11	19.23	30.77
12	42.31	30.77
13	19.23	23.08
14	53.85	46.15
15	34.62	30.77
16	34.62	26.92
Mean	38.94	34.86
SD	14.59	15.48

Appendix L

Mean difficulty score for GA and non-GA sentences for each interpreter in the Corpus Study

Interpreter	***GA Sentence***	***Non-GA Sentence***
1	0.67	1.36
2	1.6	1.2
3	2	1.18
4	1.75	1.19
5	1.55	1.07
6	1.38	1.24
7	1.67	0.92
8	1.4	1.25
9	1.17	1.32
10	1.71	1.39
11	1.7	1.33
12	1.4	1.5
13	1.71	0.68
14	2	1.19
15	3	1.15
16	2.08	0.83
17	2.2	1.26
18	2.2	0.93
19	1.5	1.45
20	1.79	0.2
21	1.83	1.33
22	2.75	1.45
23	1.54	1.82
24	1.67	0.79
25	1.65	1.75
26	2	1.53
27	1.33	2
28	1.83	1.61
29	1.91	1.46
30	1.75	1.25

Mean	1.76	1.25
Median	1.71	1.26
Sd	0.43	0.35

Appendix M

Performance Scores (%) of GA and Non-GA Interpreters (Easy Speech)

	GA-Interpreters	***Non-GA Interpreters***
1	46.66	74.84
2	62.36	56.9
3	59.76	51.74
4	69.75	55.18
5	67.71	78.36
6	52.53	67.53
7	59.82	82.38
8	48.99	59.71
Average	58.45	65.83
Median	59.79	63.62
SD	8.42	11.61

Comparison of Scores (%) of GA and non-GA Sentences (Easy Speech)

Interpreter	***GA Sentences***	***Non-GA Sentences***
1	22.00	47.08
2	50.25	63.53
3	62.50	60.00
4	75.50	70.00
5	60.78	69.08
6	28.67	53.29
7	47.06	62.35
8	28.00	53.67
Average	46.85	59.88
Median	48.66	61.18
SD	19.18	8.04

Comparison of Performance Scores (%) between GA and non-GA Interpreters (Difficult Speech)

	GA Interpreter	***Non-GA Interpreter***
1	39.87	70.13
2	46.79	53.11
3	39.87	36.57
4	51.79	61.39
5	43.66	59.67
6	33.09	37.81
7	32.75	
8	44.09	
9	36.41	
10	35.98	
Average	40.43	53.11
Median	39.87	56.39
SD	6.17	13.48

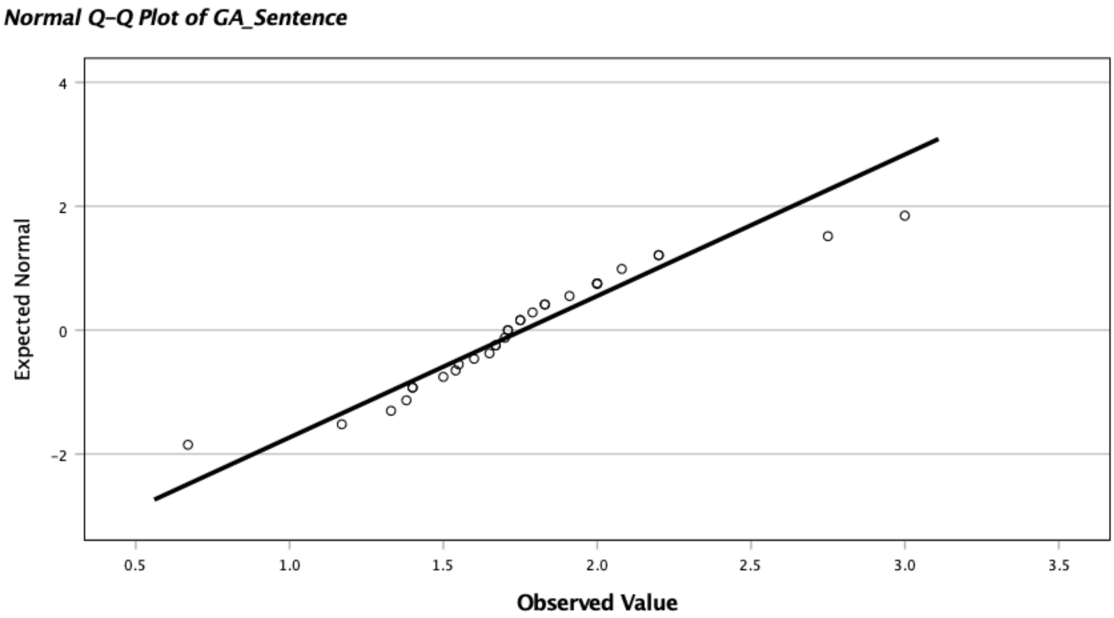
Comparison of Scores (%) of GA and Non-GA Sentences (Difficult Speech)

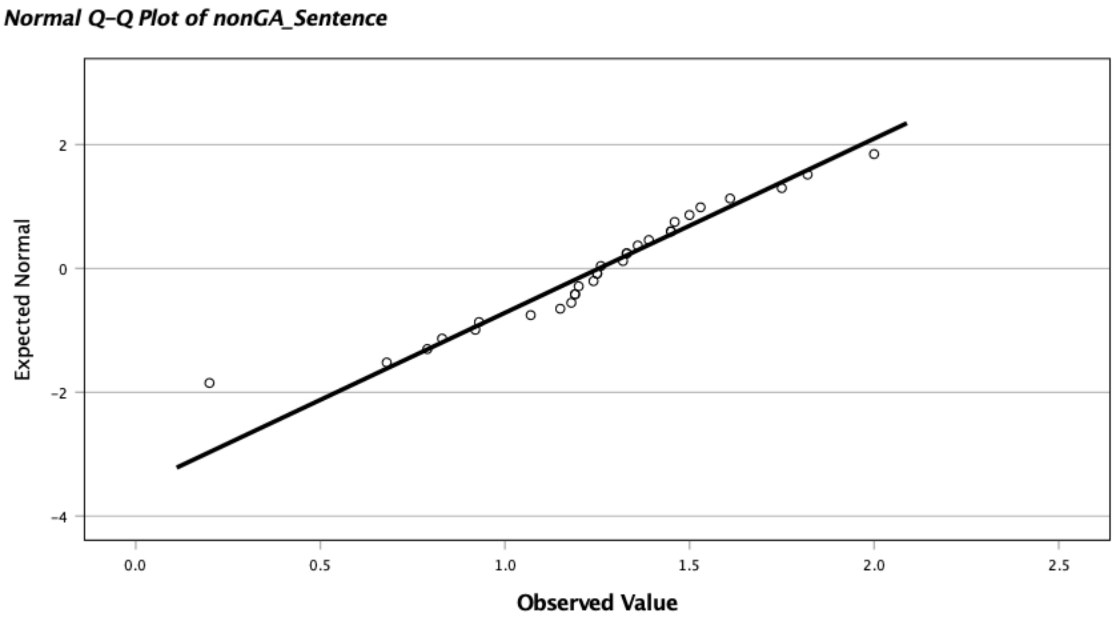
	GA Sentences	***Non-GA Sentences***
1	20.17	36.42
2	35.00	55.81
3	56.00	38.73
4	49.69	52.16
5	21.25	42.27
6	20.25	34.76
7	20.22	35.04
8	17.00	45.94
9	32.70	45.87
10	24.68	43.41
Average	29.70	43.04
Median	22.97	42.84
SD	13.57	7.14

Appendix N

Q-Q Plots

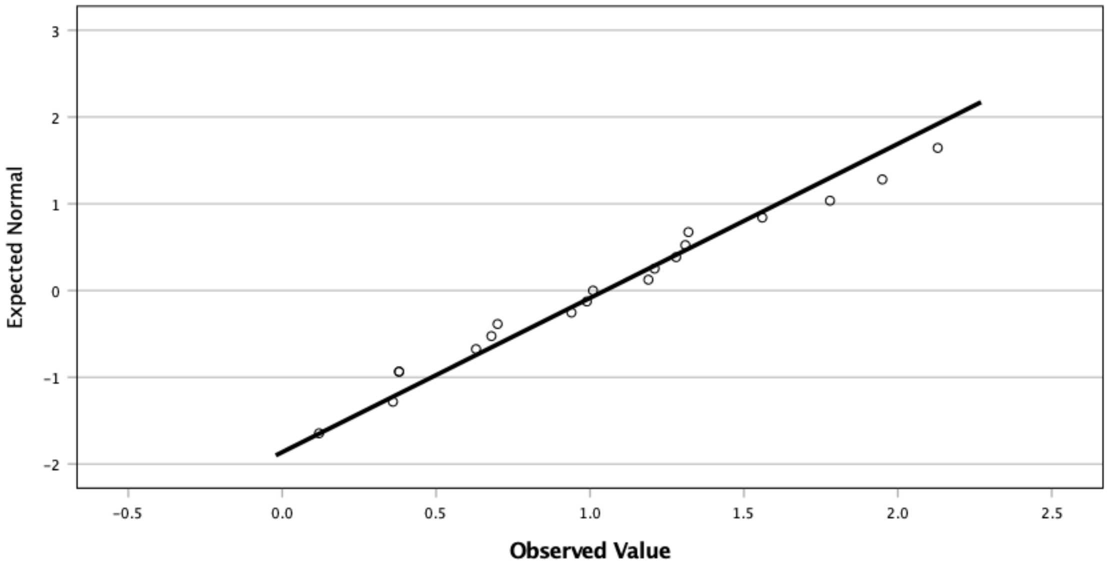
Q-Q plot for the difficulty scores of GA and non-GA sentences in corpus study



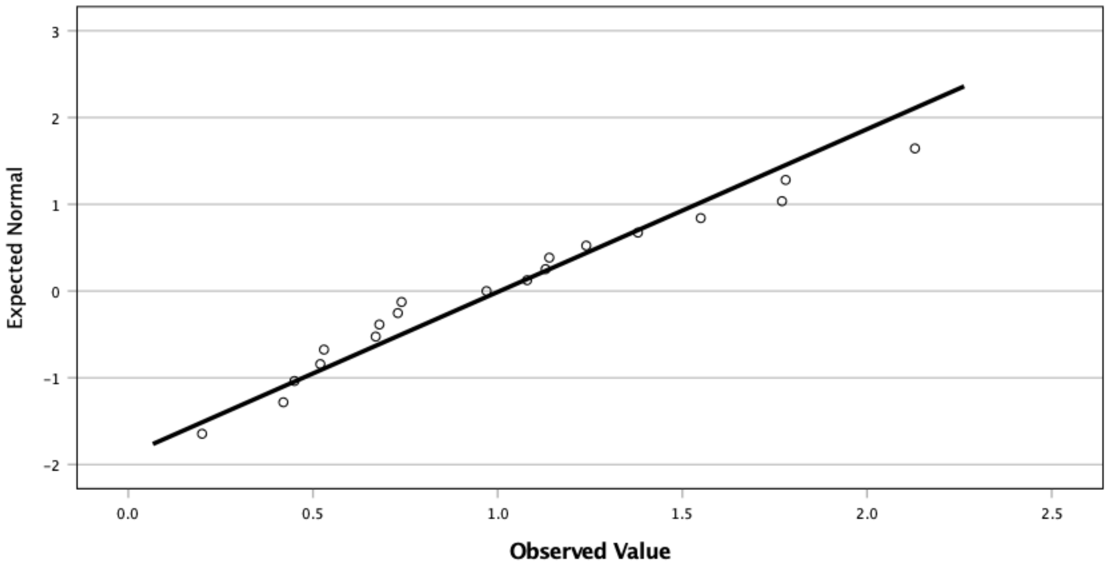


GA+NT vs GA-NT count in easy speech

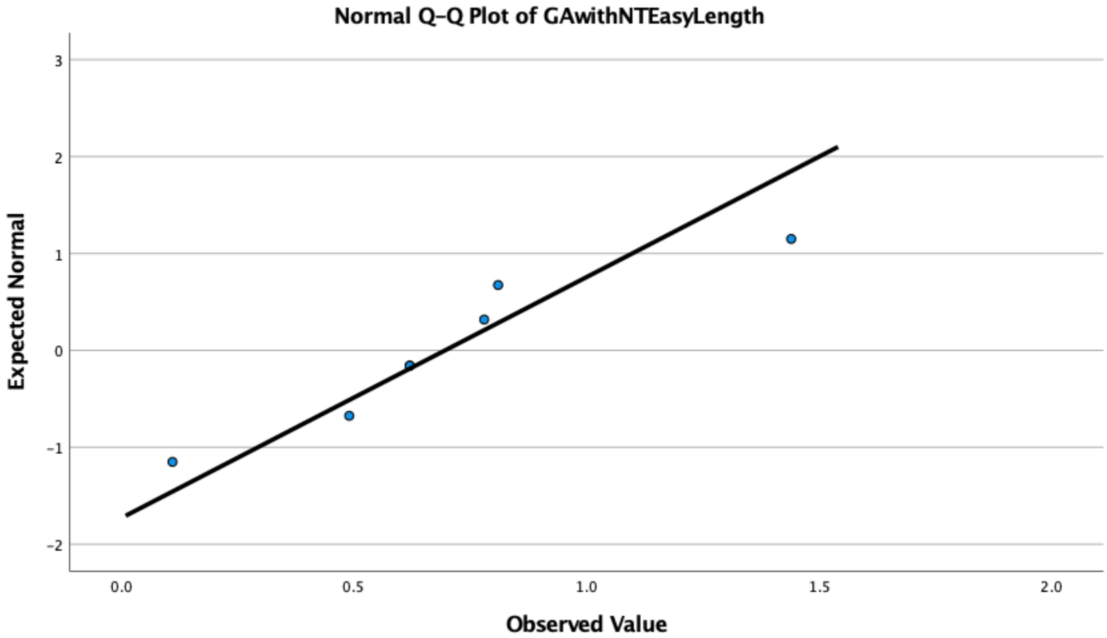
Normal Q-Q Plot of GA_NT

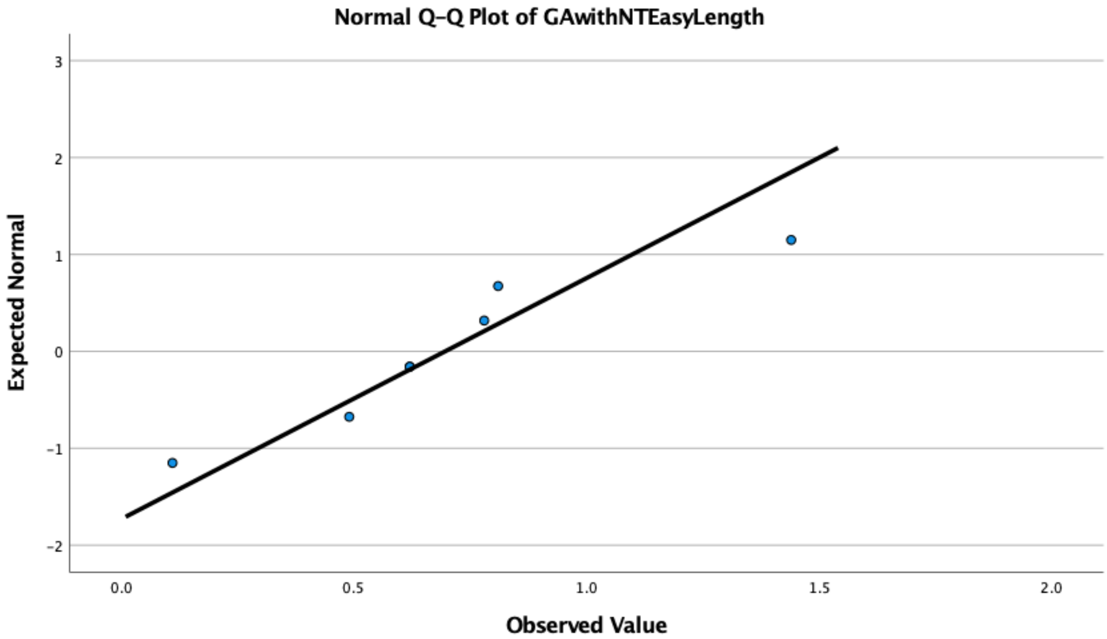


Normal Q-Q Plot of GA_without_NT

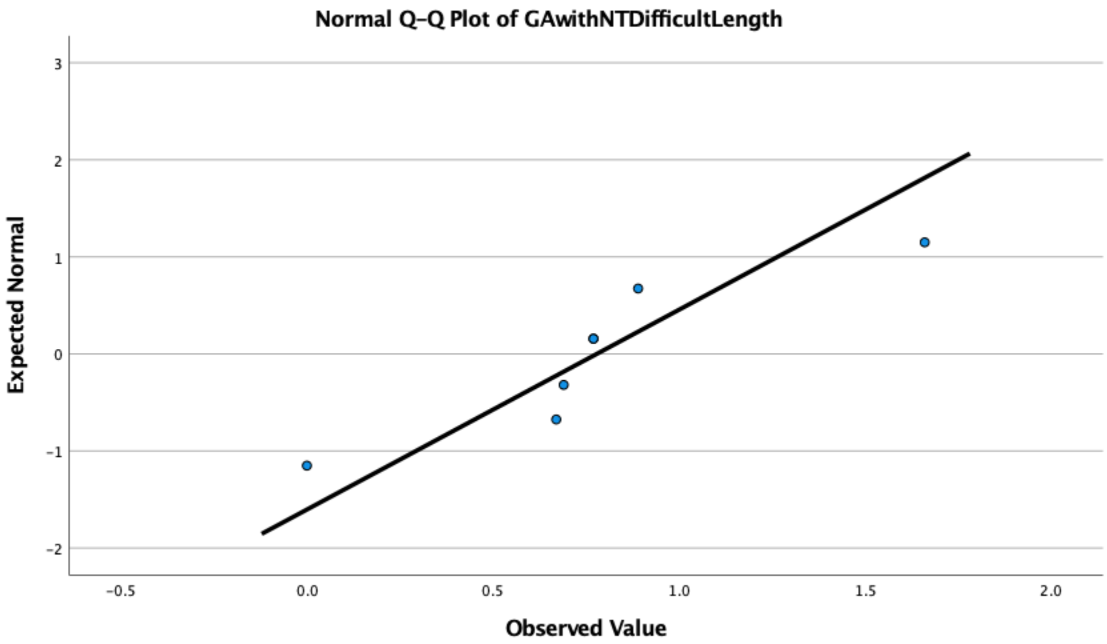


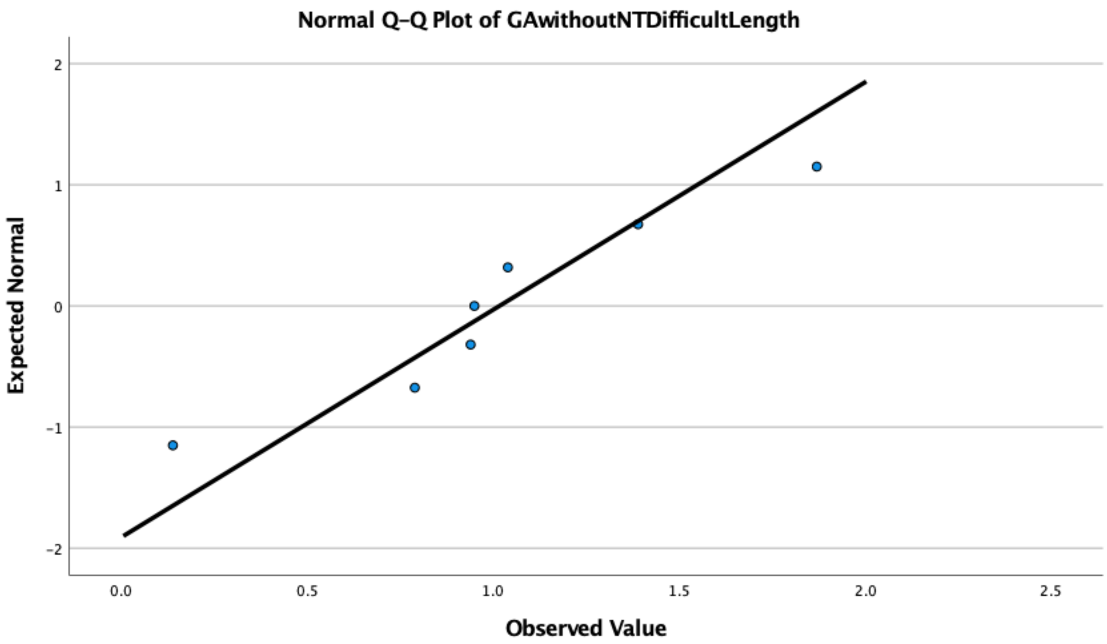
GA+NT vs GA-NT length in easy speech



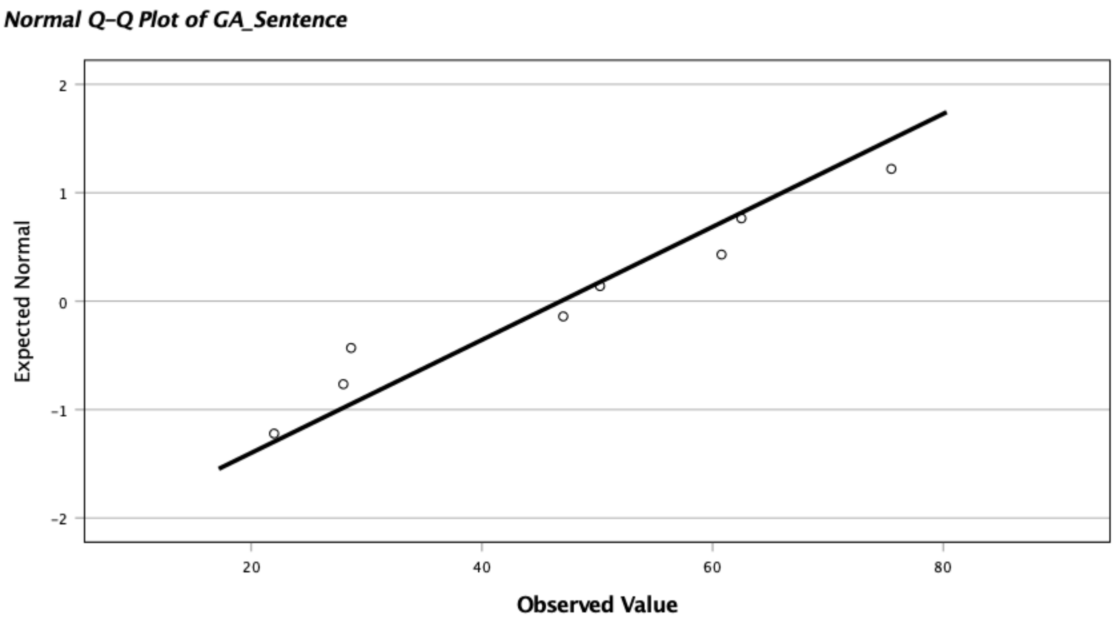


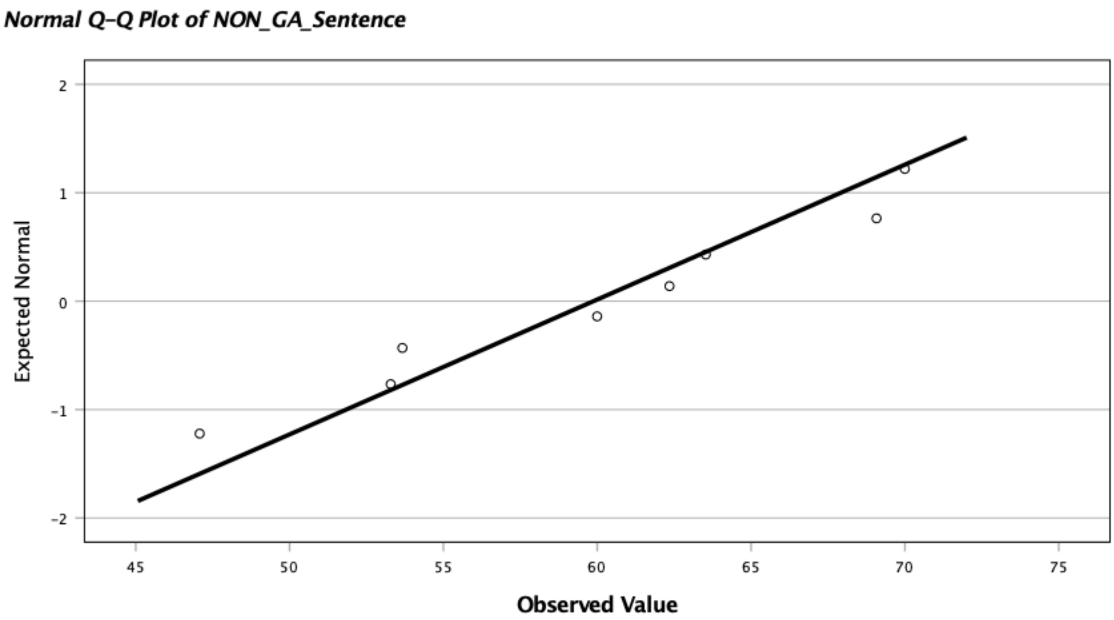
GA+NT vs GA-NT length in Difficult speech



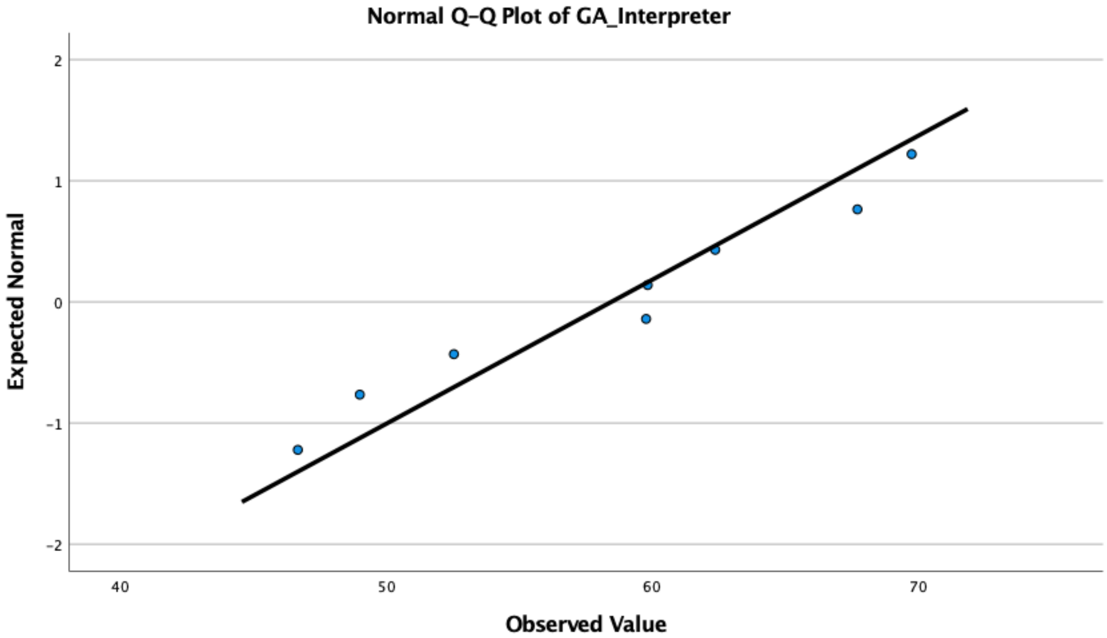


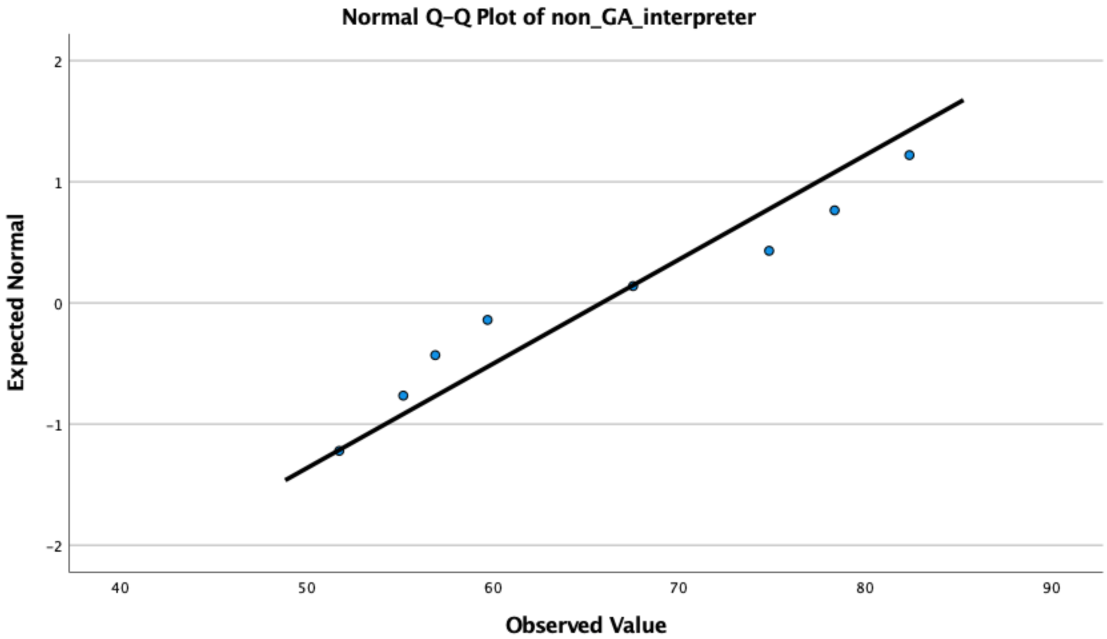
Performance score of GA and non-GA sentences in easy speech



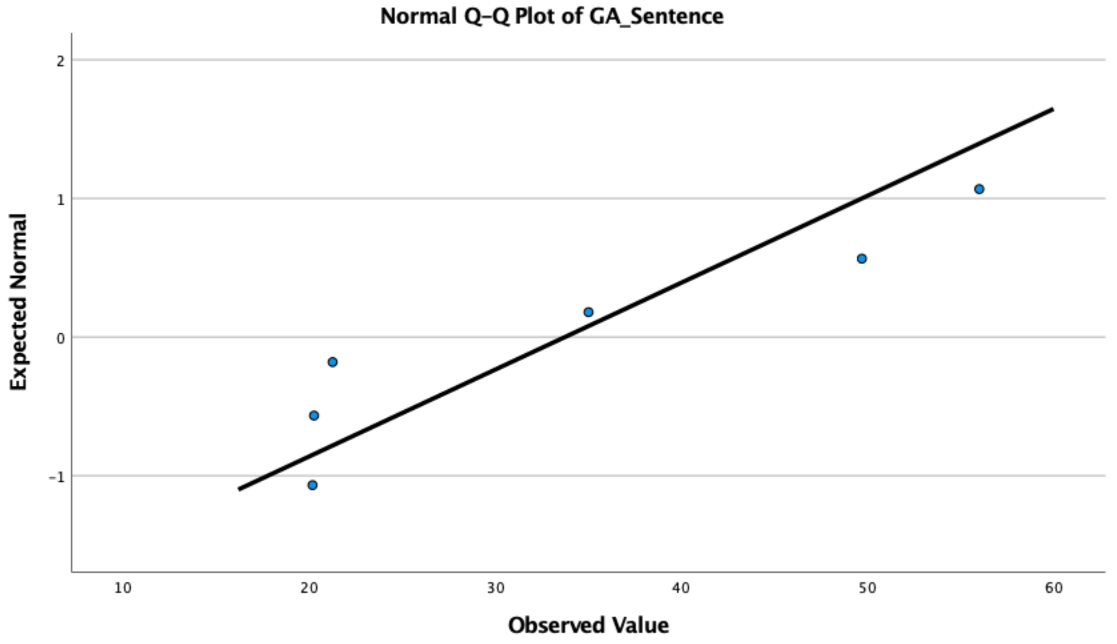


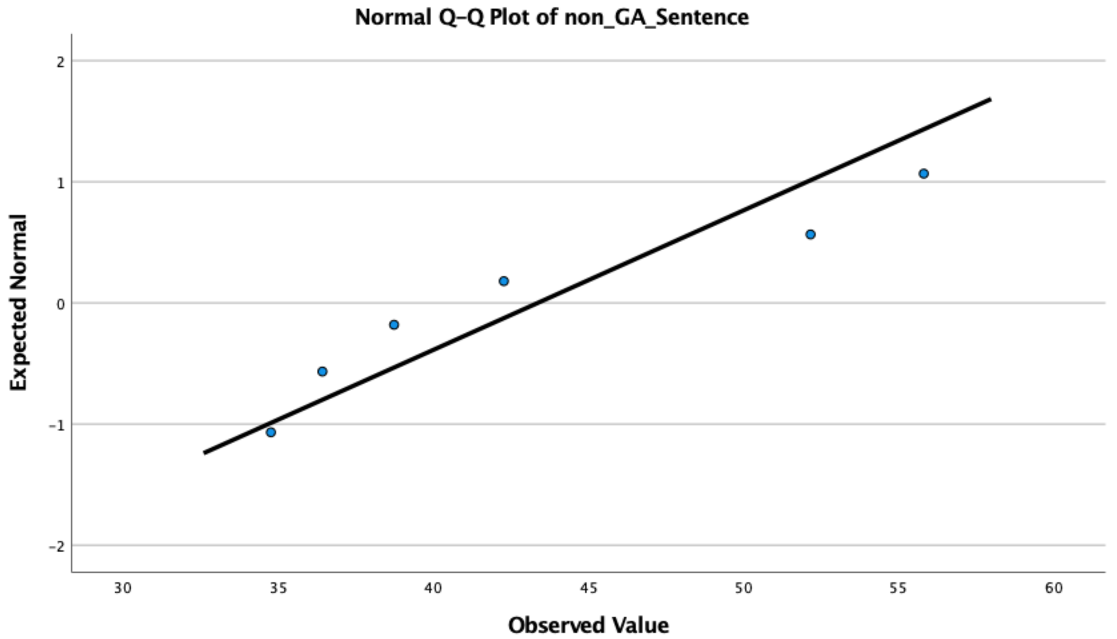
Performance score of GA and non-GA interpreters in easy speech





Performance score of GA and non-GA sentences in difficult speech





Performance score of GA and non-GA interpreters in difficult speech