# Exploring the process of note-taking and consecutive interpreting:

# A pen-eye-voice approach towards cognitive load


**Sijia Chen**

Bachelor of Arts
Master of Arts


A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy


Department of Linguistics

Faculty of Human Sciences

Macquarie University

Sydney, Australia


September 2017

# Table of contents

## List of tables

## List of figures

## List of appendices

## Abstract

Interpreting is a cognitively challenging language-processing task. Ever since it became a subject of scientific research, there has been a strong interest in finding out what happens inside the *black box* of the interpreter's mind as they perform this extraordinary task. This thesis sets out to contribute empirical evidence to elucidate the process of consecutive interpreting (CI) and note-taking, with a particular focus on cognitive load inherent in these tasks. It collects data from pen recording, eye tracking and voice recording to find answers to key questions revolving around CI and note-taking. This thesis is presented in a thesis by publication format, with its chapters (except for the introductory and concluding chapters) being stand-alone peer-reviewed journal articles.

The thesis begins with a review of the existing studies on note-taking in consecutive interpreting. It identifies the key variables of research: the choice of form (i.e., the choice between language and symbol, and the choice between abbreviation and full word), the choice of language (i.e., the choice between source and target language, and the choice between native and non-native language), and the relationship between note-taking and interpreting performance. After diagnosing two important limitations with previous studies – a lack of process research and a lack of empirical data – this review pinpoints cognitive load as a promising avenue for future investigations.

Then, the thesis presents a theoretical and methodological discussion on the construct of cognitive load in interpreting and its measurement. Borrowing from adjacent fields in which cognitive load is more systematically studied, this thesis defines cognitive load in interpreting as a multi-dimensional construct which reflects the portion of an interpreter's limited cognitive capacity devoted to performing an interpreting task in a certain environment. It introduces the categories of cognitive load measures and a series of selection criteria. Considering that previous cognitive studies mostly focus on simultaneous interpreting, this thesis introduces techniques that can be used to study cognitive load in CI.

To test the usefulness of some of the techniques proposed in the methodological discussion, a pilot study is conducted, the purpose of which is to devise a design that allows synchronised recording of pen and voice data, a combination that has been rarely applied in the field. This pilot study provides evidence that pen recording is a powerful method to tap into the process of note-taking and interpreting, thus paving the way for

the main study of this PhD project. Findings of the pilot study are also informative for the hypotheses made in the next stage of the research.

The main study of the PhD project is carried out by triangulating the methods of pen recording, eye tracking and voice recording to collect data on the process of note-taking and CI. It is found that interpreters prefer language to symbol notes and English (non-native language) to Chinese (native language) notes, regardless of the direction of interpreting. This is also the first study to visualise the activity of note-reading, showing that it proceeds in a non-linear fashion and requires significant cognitive cost. The pen and eye movement data collected in this study provide important indicators of cognitive load in note-writing, note-reading and interpreting. A combined analysis of the pen, eye and voice data shows that the note-taking choices are mainly affected by the cognitive demands, rather than the physical or temporal demands. However, the choices made by interpreters to lower the cognitive load in the first phase of CI are sometimes at the expense of interpreting performance. Furthermore, the study detects a trade-off between the cognitive costs of the two phases of CI.

Understanding the nature of the cognitive processes involved in interpreting is not only beneficial to the field itself – to inform interpreter education, testing and continuing professional development – but also more generally enriches our understanding of bilingual language processing and human cognition. The methodological and empirical findings of the thesis contribute to that effort and outline possible avenues for future research.

## Declaration

I certify that the work in this thesis is the result of my own research and that the work has not been submitted for a higher degree to any other university or institution. I certify that sources of information used and the extent to which the work of others has been utilised are indicated in the thesis.

The research was approved by Macquarie University Ethics Committee (Human Research) (see Appendix B) and conducted in accordance with the guideline stipulated.

Sijia Chen

September 2017

## Acknowledgements

I am immensely grateful to my principal supervisor Associate Professor Jan-Louis Kruger. Words can't even begin to describe how much help and support Jan-Louis has provided me through the entire PhD candidature. His mentorship and high standard of scholarship prepared me well for the various challenges that came along the journey. I couldn't ask for a better supervisor. The fulfilment of this PhD project would not have been possible without him.

I would like to express my sincerest gratitude to my co-supervisor Dr Helen Slatyer and my adjunct supervisor Dr Stephen Doherty. Without Helen, I wouldn't even have been able to start my PhD at Macquarie University. She helped me get through the hardest times, especially the beginning stage which was filled with struggles. Stephen offered me generous and invaluable support in the stressful moments of my final year. His guidance, insights and feedback are indispensable to the successful completion of this PhD research.

I would also like to thank Professor Jing Chen, who supervised my undergraduate and postgraduate studies at Xiamen University and introduced me to the field of interpreting research.

My sincere gratitude also goes to Professor Gile Daniel for providing the Conference Interpreting Research Information Network Bulletins (CIRIN Bulletins), one of the key databases for the literature review of this thesis. I am deeply touched by his patience and generosity in sharing his insights with a young researcher like me.

I am grateful to all my colleagues and friends for their kind help and mutual encouragement. A very special thank you goes to Dr Chao Han for inspiring and encouraging me both as a friend and as a peer researcher, and for providing critical feedback on the thesis chapters.

I also owe my gratitude to all the participants, speakers, technicians, administrators, and all other university staff who have contributed their precious time to help me carry out the study.

I greatly appreciate the valuable feedback from journal editors and anonymous reviewers on my published and submitted papers. I also want to thank the reviewers of this PhD thesis for dedicating their time and effort.

My sincere gratitude goes out to the China Scholarship Council and Macquarie University for co-funding the PhD research.

To my husband Dong Liu: thank you for your boundless love, enormous patience, and constant encouragement. Thank you for helping create little master Yixuan Liu in the middle of my PhD journey. I love you.

To my baby boy: your smile can easily chase away all my anxiety, even in the toughest times.

致父母：亲爱的爸爸妈妈，你们的爱，伴我成长，给我力量。我爱你们。

# Chapter 1 Introduction

This opening chapter introduces the literature and theoretical underpinnings of the thesis, points out the limitations and gaps in previous studies, and articulates the aims and questions of the research. It introduces the design of the project, with paticular attention paid to the methodological triangulation of pen recording, eye tracking and voice recording to study interpreting. Presented in a thesis by publication format, this thesis consists of an introductory chapter providing the background and framework for the study, followed by a series of individual research articles linked together into one overall argument, and a concluding chapter providing the overarching conclusions of the project as a whole.

## 1.1 Literature and theoretical underpinnings

### 1.1.1 Investigating the process and cognitive aspects of interpreting

Interpreting is an intriguing, challenging, and complex language processing task. Ever since interpreting research became established as a field of study in its own right in the mid-1970s (Pöchhacker, 2004, p. 81), there has been a strong interest in uncovering what is happening in interpreters' minds while they perform this extraordinary task. Researchers with a background in psychology have attempted to shed light on how the human mind processes language under severe stress and while engaging in heavy multi-tasking by investigating the cognitive processes in interpreting (e.g., Barik, 1973; Christoffels, 2004; Christoffels & De Groot, 2004, 2005; De Groot, 1997; Gerver, 1974a, 1974b, 1976; Goldman-Eisler, 1972; Köpke & Signorelli, 2012). Researchers from within the field of interpreting, in turn, approached the topic from an inter-disciplinary perspective that benefits from the theoretical and empirical findings in the cognitive sciences (e.g., Lambert, 1988; Moser-Mercer, 1997; Seeber, 2011, 2013; Shlesinger, 2000).

However, most of the process-oriented research approaching interpreting from a cognitive perspective focuses on simultaneous interpreting (SI), while consecutive interpreting (CI) is often neglected. CI was the first form of interpreting used at international conferences and dominated the market in the first half of the 20th century. It gradually gave way to SI, which was made possible by the development of electronic equipment, in multilateral and multilingual conference settings. However, CI remains the preferred mode in the context of "bilateral interactions with only two languages involved and in settings where confidentiality, intimacy and directness of interaction are given priority over time efficiency", such as high-level diplomatic encounters, business negotiations, ceremonial speeches and press conferences (Dam, 2010, p. 76). CI remains an important component in most interpreter training programmes. Its significance is manifested in the large quantities of master's theses on the subject[1]. Even in places where the market is largely dominated by SI, training in CI is believed to be a good way of preparing students for SI (Gile, 2001). Furthermore, CI is frequently introduced to language students as a way of reinforcing language skills (e.g. Henderson, 1976; Hill, 1979; Paneth, 1984).

Given the important role CI plays in the above contexts, there exists a considerable limitation in the literature in that process-oriented cognitive investigations have rarely been carried out on CI. CI is an interesting activity from both a cognitive and a linguistic point of view. Similar to SI, it requires a high level of bilingual language processing and challenges the interpreter's cognitive system by requiring multi-tasking under strict time constraints. But CI also introduces a new challenge: note-taking[2]. In addition to listening to the source speech and producing a target speech, CI requires the interpreter to perform the tasks of note-writing and note-reading. In Phase I of CI, interpreters listen to and analyse the source speech, keep parts of the speech in their working memory, and write down notes. In Phase II, interpreters read back their notes, retrieve information from their working memory, and produce a target speech. Both phases depend heavily on note-taking – this unique and distinctive feature of CI.

### 1.1.2 Note-taking in CI

Note-taking has been a topic of interest in interpreting research for over half a century (see Chapter 2). The well-developed volume of literature on consecutive note-taking started with a series of books and articles introducing various note-taking systems and principles. They were published in different languages, each generating a profound influence in its own country and some even reached beyond (e.g., Allioni, 1989; Becker, 1972; Gillies, 2005; Gran, 1982; Ilg, 1988; Kirchhoff, 1979; Matyssek, 1989; Rozan, 1956/2002). Recommendations were made on such skills as noting the idea and not the

[1] Interested readers can find the theses reported in various issues of the Conference Interpreting Research Information Network Bulletin (CIRIN Bulletin) at www.cirinandgile.com.

[2] In this thesis, consecutive interpreting refers to long consecutive where systematic note-taking is used.

word, how to use symbols, how to use abbreviations, and how to note links, negations, and emphasis.

With such well-developed note-taking systems, it would seem that once students are made aware of the systems and practice accordingly, note-taking should pose few problems. However, when it comes to teaching and learning note-taking skills, both teachers and students find it challenging. Various studies (e.g., Alexieva, 1994; Gile, 1991) identified that note-taking diverted students' attention and even led to a degradation in interpreting performance.

Researchers who have approached the topic form cognitive and linguistic perspectives (Kirchhoff, 1979; Kohn & Albl-Mikasa, 2002; Seleskovitch, 1975) found that there was a concurrent storage of information in notes and in memory, and a competition for cognitive resources between note-taking and other activities in the interpreting process. This has motivated subsequent research to target more specific note-taking features and to examine them empirically.

Some of the most important variables investigated are: the choice of form (e.g., Dai & Xu, 2007; Dam, 2004a), the choice of language (e.g., Abuín González, 2012; Dai & Xu, 2007; Dam, 2004b; Szabó, 2006), and the relation between note-taking and interpreting performance (e.g., Cardoen, 2013; Dai & Xu, 2007; Dam, 2007; Dam, Engberg, & Schjoldager, 2005). The choice of form refers to the choice between language and symbol, and the choice between abbreviation and full word. The choice of language refers to the choice between source and target language, and the choice between native and non-native language. Despite a couple of general trends, such as a preference for language over symbol and a source language dominance in the notes taken by student interpreters, the studies have reported inconsistent findings (see Chapter 2).

The inconsistencies are potentially related to some shared limitations of these studies. Of the limited empirical data that have been collected, a large portion has been collected from students (varying in their stages of study and maturity of competence), making the findings difficult to generalise (Gile, 2009, p. 179). Furthermore, many studies only investigate one interpreting direction so the results are hard to compare. More importantly, most of the studies are product oriented, which means that they only look at the final product of note-taking (the notes produced), without an in-depth analysis of the interpreting process. Last but not least, the research efforts are somewhat scattered, without an overarching framework to pull them together.

A possible solution is to examine the topic from a cognitive perspective. As Gile (2009, p. 178) points out, “note-taking is an area in which the concept of processing capacity can be useful.” Viewed from a cognitive perspective, all discussions on note-taking could boil down to one fundamental question: how to reduce the cognitive load of note-taking while maintaining the efficiency of notes. If cognitive load can be measured while interpreters perform note-taking and CI, some fundamental principles underlying the note-taking choices might be unveiled. It is possible, for example, that the different observations reported by previous studies might not be controversies, but rather converging evidence in proving that interpreters make choices according to their own characteristics to reduce cognitive load.

The current research attempts to revisit the topic of note-taking and CI and address some of the existing limitations by (1) using professional interpreters as participants; (2) investigating both directions of interpreting; (3) combining product analysis with the process research methods of pen recording, eye tracking and voice recording; and (4) performing analyses and illustrations under a common cognitive framework and focusing on cognitive load in the process.

### 1.1.3 The construct of cognitive load in interpreting and its measurement

The construct of cognitive load was not new when it was introduced to the field of interpreting in the 1980s. Similar constructs had already been investigated, with two of the most relevant ones being mental workload in human factors research and cognitive load in Cognitive Load Theory. Human factors research centres on how humans accomplish tasks in the context of human-machine system operation and how different variables affect that accomplishment (Meister, 1989). Cognitive Load Theory is a theory of learning, focusing on how instruction affects the load on learners’ cognitive systems (Sweller, Ayres, & Kalyuga, 2011). The two fields define their central constructs using different terms (see O'Donnell and Eggemeier (1986) for mental workload and Paas and Van Merriënboer (1994) for cognitive load), but both constructs capture the interactions between a task with specific characteristics and a human with limited cognitive capacity. In that sense, both constructs are relevant and both fields serve as good references for the definition and illustration of cognitive load in interpreting. Borrowing from these two fields, this research therefore builds a theoretical and methodological framework for defining and measuring cognitive load in the field of interpreting (see Chapter 3).

Cognitive load in interpreting has been operationalised as a multi-dimensional construct depending on the interactions between two groups of variables: the task/environmental characteristics and the interpreter characteristics. Since cognitive load is a theoretical construct, it cannot be observed or measured directly. Its measurement relies on an arrangement of surrogates that are indicative of cognitive load. Moreover, cognitive load is a multi-dimensional construct, so any single measure cannot provide a comprehensive picture. Due to such challenges, only a few pioneering studies have investigated cognitive load in interpreting (see Chapter 3). These studies usually apply only one technique to examine the cognitive load, and are almost exclusively on SI. This research directly contributes to that effort by using a combination of methods, including pen recording, eye tracking and voice recording, to explore the cognitive load in CI.

## 1.2 Research scope, research purpose and research questions

The scope of this process-oriented cognitive research on interpreting is limited in four respects. First, the study focuses only on CI – a spoken-language interpreting in the consecutive mode – and not on other interpreting modalities. This is mainly due to a lack of research on the processing and cognitive aspects of CI and the fact that the author was only trained and practiced in spoken-language (as opposed to sign-language) interpreting. Second, the research only involves the language pair of Chinese and English. This is because the author only has the linguistic prerequisites to study and analyse this particular language pair at the required level. Third, the project only recruited professional interpreters to participate in the experiment. This is to address the problem that most previous studies on note-taking and CI only involved student interpreters. Despite its necessarily limited scope, this study is expected to make methodological contributions to and generate meaningful findings for future research involving other interpreting modalities, language pairs, and interpreter types.

The research has three main purposes. First, it aims to make a methodological contribution to bilingual language processing research (especially translation and interpreting research) by triangulating the methods of pen recording, eye tracking, and voice recording. Second, it seeks insights into the interpreter’s *black box* by examining the cognitive processing and cognitive load in interpreting. Third, through observing and analysing how professional interpreters take notes in CI, the study searches for empirically-based recommendations for interpreter education and continued professional development.

The PhD study sets out to answer three main research questions (RQs) corresponding to the key variables in consecutive note-taking identified in the literature.

*RQ 1: What are the preferred note-taking choices of professional interpreters in terms of form and language?*

This RQ subsumes two sub-questions, including:

*RQ 1.1 What do interpreters prefer when choosing the form of note-taking: language or symbol; abbreviation or full word?*

*RQ 1.2: What do interpreters prefer when choosing the language of note-taking: source or target language; native or non-native language?*

After identifying the patterns of note-taking choices, a second step is to examine the relation between these note-taking choices and cognitive load in CI.

*RQ 2: What is the relationship between note-taking choices and cognitive load in CI?*

More specifically, the study is interested in how the choices made by interpreters during Phase I of CI affect the level of cognitive load in Phase II. So RQ2 also consists of two sub-questions:

*RQ 2.1: What is the relationship between the note-taking choices and cognitive load in Phase I (the listening and note-writing phase) of CI?*

*RQ 2.2: What is the relationship between the note-taking choices and cognitive load in Phase II (the note-reading and production phase) of CI?*

The third research question concerns a central issue in interpreting research, namely the quality of performance, and its relationship with note-taking. The aim is to see whether specific note-taking choices help or hinder the interpreting performance.

*RQ 3: What is the relationship between note-taking and interpreting performance?*

## 1.3 Research design

This PhD research used an exploratory design to gain insights into the process of note-taking and CI. In particular, it triangulated the methods of pen recording, eye tracking and voice recording to collect data on cognitive processing and cognitive load in interpreting. To make the data more generalizable, professional interpreters rather than student interpreters (whose interpreting competence are greatly varied and not yet mature) were recruited. Two CI tasks covering both directions of interpreting (between Chinese and English) were involved to account for both the source/target language status and the native/non-native language status. The order of the two CI tasks was randomised to eliminate the impact brought by task order. A retrospection was designed to collect additional qualitative data following the completion of the two tasks. The note-writing process was recorded via pen recording; the note-reading process was recorded via eye tracking; the interpreting process and retrospection were recorded via voice recording.

### 1.3.1 Pen recording

The apparatus used for pen recording was the *Cintiq 13HD* (a 13-inch LCD tablet with a resolution set at 1366×768 pixels) and the *Wacom Pro Pen*. The system was chosen because it targets graphic designers who have very high requirements in terms of the precise control of the pen on the tablet surface. It is ergonomically designed to mimic natural writing and painting. Another reason for choosing this system is because it is compatible with the *Eye and Pen* software[3], one of the core software products powering the experiment. The software piloted a laptop computer which was linked to the pen recording apparatus. The software carried out three tasks: controlling the experiment, collecting the pen data, and processing the pen data.

*Controlling the experiment*. The experiment and its procedures were programmed into the software, which then controlled the progress of the experiment and interacted with the participant. For example, in Phase I of CI, when finishing one page of note-taking, the participant could use the pen to click on a button displayed on the tablet screen called "New Page" (Figure 1.1) and the software would create a new blank page for note-taking. The participant could use as many pages as needed. When the listening and note-writing phase was finished, the participant only needed to click a button called "Begin Interpreting" (Figure 1.1) and the software would automatically turn to the first page of notes written by the participant. Then the participant could read back the notes and produce a target speech. In this phase, new buttons such as "Turn Page" (which turns to the next page of written notes) and "Next part" (which plays the next segment of the source speech) would appear on the screen and the participant could interact with the software to navigate through the pages of written notes. The tablet screen would only react to the tip of the digital pen, so the participant could write as naturally as

[3] More detailed information about the software can be found on http://eyeandpen.net/en/.

possible and did not need to worry about triggering any buttons by touching the screen with their hands.


Figure 1.1 A screenshot of the tablet in the recording mode

*Collecting the pen data*. The software collected the spatial and temporal data about the pen as it moved across the tablet surface. For example, data was recorded for each pen stroke in terms of the distance (how far the pen travelled across the surface), duration (for how long the pen was in touch with the tablet), and speed (how fast the pen was moving). Spatial data was reported in centimetres and temporal data was reported in milliseconds. The software also kept a session log for each trial, documenting the time every action took place during the recording (e.g., the source speech segment started playing, the participant started writing, etc.). This function was crucial for the calculation of an important cognitive load indicator, the ear-pen span, which is the time span between the moment a speech unit is heard and the moment it is written down in notes (see Chapter 4 and Chapter 5).

*Processing the pen data*. The software has many functions for displaying and analysing the recorded pen data (Figure 1.2 is a screenshot of the software in the analysis mode). The most useful function for this study is the “Word separation” tool, which semi-automatically separates the written texts into words (in this study’s case, note units). Although manual work was required to correct the separations, this function allowed very accurate data to be reported for each individual note unit (e.g., start and end time, duration, distance, speed, etc.). Labels could be created for each note unit so that qualitative data could be added to each note and exported for further analysis. For example, for note unit no.13 (see bottom left of Figure 1.2), texts 1 to 6 documented the form and language of the note unit as well as its content, meaning, and corresponding source speech unit. The labels indicated that this note unit was language ("L" in Text 1), in English ("E" in Text 2), and an abbreviation ("A" in Text 3). It contained three letters "svs" (Text 4), meaning "services" (Text 5) and corresponding to the word "services" (Text 6) in the source speech. In this way, the exported file contained both quantitative and qualitative data (Figure 1.3).

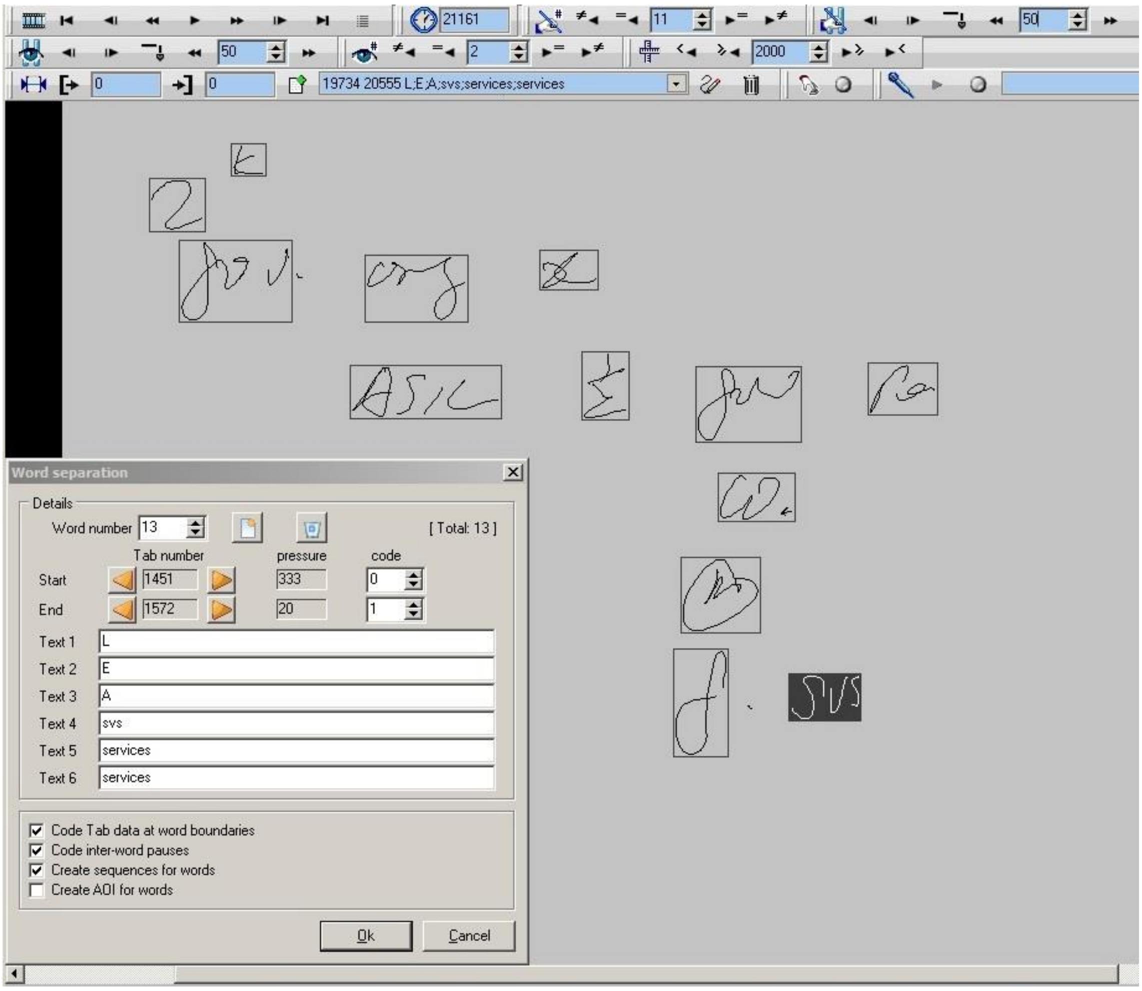


Figure 1.2 A sample screenshot of the Eye and Pen software in the analysis mode

| (d=centimeter; t=millisecond; v=centimeter/second) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SeqNum | Label 1 | Label 2 | Lable 3 | Content | Meaning | Source unit | Begin time | End time | nb_pts | distance | duration | speed |
| 1 | N | | | 2 | 2 | 2 | 2538 | 2860 | 58 | 2.4386 | 322 | 7.5731 |
| 2 | L | E | A | k | know | know | 3471 | 3827 | 53 | 1.9384 | 356 | 5.445 |
| 3 | L | E | A | gov | governme | governmen | 4116 | 5153 | 175 | 6.919 | 1037 | 6.6721 |
| 4 | L | E | A | org | organisatio | organisatior | 5765 | 6565 | 144 | 5.34 | 800 | 6.6751 |
| 5 | S | | | )( | deal with | deal with | 7281 | 7776 | 89 | 3.3085 | 495 | 6.6839 |
| 6 | L | E | A | ASIC | ASIC | the Australi | 8342 | 9652 | 185 | 9.5906 | 1310 | 7.3211 |
| 7 | L | C | A | 立 | 独立的 | independer | 13401 | 14129 | 110 | 3.8353 | 728 | 5.2683 |
| 8 | L | E | A | gov | governme | governmen | 14441 | 15313 | 157 | 5.7602 | 872 | 6.6057 |
| 9 | L | E | A | bo | body | body | 15524 | 16128 | 105 | 3.8048 | 604 | 6.2994 |
| 10 | L | E | A | co | corporatio | corporation | 16650 | 17528 | 129 | 4.7633 | 878 | 5.4252 |
| 11 | S | | | mO | market | markets | 17900 | 18673 | 139 | 6.6106 | 773 | 8.5519 |
| 12 | L | E | A | f | financial | financial | 18973 | 19356 | 69 | 4.3474 | 383 | 11.351 |
| 13 | L | E | A | svs | services | services | 19734 | 20555 | 122 | 4.6932 | 821 | 5.7164 |

Figure 1.3 A sample data output of pen recording

### 1.3.2 Eye tracking

There were a few prerequisites for selecting the type of eye tracker to be used in the study. First, the eye tracker needed to allow the interpreter to speak freely, thus eliminating the use of eye trackers that require chin rests. Second, the eye tracker needed to be usable in a handwriting situation. In particular, the eye camera(s) could not be masked by the participant's forearms in movement. Head-mounted eye trackers could meet the first two requirements. Third, for the comfort of the participant and ecological validity of the experiment, a light-weight eye tracker that could be attached to the participant easily was preferable.

The eye tracker used in this study was the *SensoMotoric Instruments (SMI) Eye Tracking Glasses 2 (ETG2)*. It is a light-weight (47 g), head-mounted eye tracker in the shape of a pair of glasses. The eye tracker uses dark pupil tracking. It has a tracking accuracy of .5° over all distances and a sampling rate of 60 Hz. The eye tracker has a built-in high-definition camera for scene recording. This camera recorded both the video and the audio during the entire note-taking and interpreting process. The SMI software *iView ETG* and *BeGaze* were used with default settings for eye data recording and analysis respectively. The experiment took place in a sound-proof studio with constant artificial illumination to avoid any distractions or disruption to the recording of eye data.

### 1.3.3 Voice recording

Voice recording was used for several different purposes in this study. First and foremost, the interpreting performance was recorded. The audio recordings were later transcribed and provided to a group of raters for evaluation (see Chapter 4 and Chapter 5). This generated performance scores used for exploring the relationship between note-taking and interpreting performance. Second, voice recording was used during cued retrospection. Immediately after the interpreting tasks, the participants were provided with their notes for cued retrospection. They were asked to provide as much information as they could remember about the note-taking process, including but not limited to: what each note unit was; what it stood for; whether it was symbol or language, and if language, whether it was abbreviation or full word, Chinese or English. This is an important step because note-taking in CI is highly individualised, and the handwriting of interpreters could sometimes be difficult for others to decipher. Third, the source speech audio files were used together with the session logs kept by the Eye and Pen software to calculate the ear-pen span, an important indicator of cognitive load used in this study (see Chapter 4 and Chapter 5).

### 1.3.4 Participants, tasks and procedures

A total of 26 professional interpreters were recruited using purposive sampling. Ethics approval was obtained for conducting the research (see Appendix B). Table 1.1 summarises the criteria used to recruit the participants. Demographic questionnaire (Appendix C) results showed that most of the participants had a postgraduate degree in interpreting (65%); some had an interpreting diploma (15%); some attended an intensive interpreting training course (15%); and one was self-trained (4%). The age of the participants averaged at 36.4 years; their working experience (years of working as either full-time or part-time interpreters) averaged at 7.4 years. The participants worked most frequently in Australia (with only two exceptions who worked more frequently in China).

Table 1.1 Criteria for recruiting participants

| | |
|---|---|
| Accreditation | *Professional Interpreter* level accreditation from Australia's National Accreditation Authority for Translators and Interpreters (NAATI) |
| Working language | Mandarin Chinese (native language) and English (non-native language) |
| Experience | A minimum of two years of experience working as part-time or full-time interpreters;<br>If working as a part-time interpreter, the other jobs have a bilingual feature (e.g., interpreter trainer);<br>Experience in working in the consecutive mode |

Two CI tasks (see Appendix E) were carefully created through a series of procedures to control for variance (see Chapter 4). The experiment took place in four main procedures (see Chapter 6 for a detailed procedure): practice, task performance, retrospection and post-experiment questionnaire. The practice session was designed to familiarise the participants with the experimental procedures and the apparatus, especially the digital pen and the eye tracker. The task performance session involved two CI tasks, the order of which was randomised. Rest was allowed between tasks if needed. The retrospection session was cued by the written notes and participants were instructed to recall whatever they could remember about the note-taking process. This was mainly designed to help the researcher accurately identify the note units and to collect additional qualitative data for the interpretation of the results. The questionnaire was designed to collect such information as the participants' familiarity with the task topics, how they felt about using the digital pen and the eye tracker, and other feedback about the experiment.

### 1.3.5 Data and analysis

The main sources of data collected in this study and their corresponding RQs and chapters are summarised in Table 1.2. Wherever applicable, the data were standardised using a note unit as the unit of analysis. For example, if the number of Chinese notes written by a participant is *n*, then the ear-pen span of Chinese notes ($EPS_C$) of that participant is calculated as:

$$EPS_C = \frac{1}{n} \times (EPS_1 + EPS_2 + \cdots + EPS_n)$$

Paired-samples t-tests were conducted to compare between the note-taking choices in different forms (language vs. symbol; abbreviation vs. full word) and languages (Chinese vs. English). The Pearson's correlation was used to explore the relation between note-taking and interpreting performance. All statistical analyses were performed by running the IBM SPSS Statistics 22. Two-tailed $p$ values less than 0.05 were considered to be statistically significant. Cohen's $d$ (the difference between the means divided by the pooled standard deviation) was used to indicate the effect sizes, which were classified as small ($d$ = 0.2), medium ($d$ = 0.5), and large ($d$ = 0.8).

Table 1.2 Data collected in the study

| **Source** | **Data** | **RQ(s)** | **Chapters** |
|---|---|---|---|
| Pen recording | All written note units (including the form, language, content, meaning, and corresponding source speech unit) | RQ 1.1 and 1.2 | Chapter 4 and 5 |
| | The distance, duration, speed and ear-pen span of all note units | RQ 2.1 | |
| Eye tracking | Eye movement measures such as regression rate, first fixation duration, first-pass dwell time, second-pass dwell time, total dwell time, number of fixations, number of revisits, average fixation duration, and skip rate | RQ 2.2 | Chapter 6 |
| Voice recording | Interpreting performance and audio of retrospection | RQ 3 | Chapter 4, 5 and 6 |

## 1.4 Outline of the thesis

This thesis is written in a "thesis by publication" format, an encouraged and preferred approach at Macquarie University for Higher Degree Research candidates. According to the university's guideline (see Appendix A), a thesis by publication may include relevant papers which have been published, accepted, submitted or prepared for publication during the candidature. The papers need to be sufficient to support the important findings from the research and presented in a logical and coherent way. Most theses by publication have between 2 and 8 papers, each forming a chapter of the thesis.

This thesis consists of seven chapters. Chapter 1 is the introductory chapter and Chapter 7 is the concluding chapter. Chapters 2 to 6 are self-contained journal articles. Chapters 2, 3, and 4 have been published in peer-reviewed journals; Chapter 5 is currently under peer review and Chapter 6 has been submitted to a peer-reviewed journal. Relevant publication details are specified in a footnote at the beginning of each chapter. The links between the article-based chapters are illustrated in Figure 1.4.

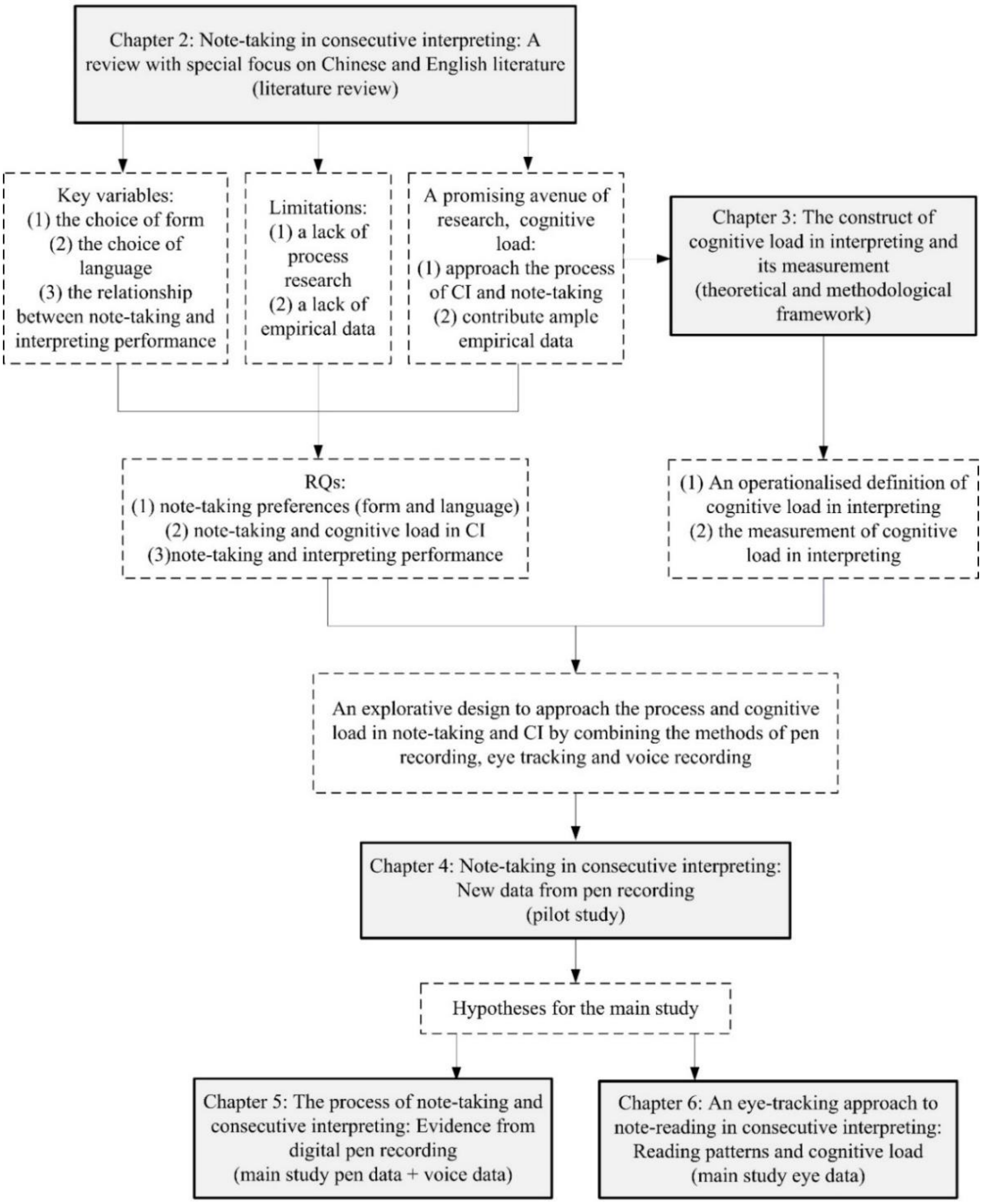


Figure 1.4 Linkages between the article-based chapters in this thesis by publication

Chapter 1 gives an overview of the literature and theoretical underpinnings of the study, states the scope, purpose and questions of research, and introduces the design of the project.

Chapter 2 is a critical review of the studies that have been carried out on consecutive note-taking. It sets the literature background for the thesis and identifies the most important variables on the topic. Two major limitations of previous studies are identified as a lack of process research and a lack of empirical data. Cognitive load is established as a promising avenue for future research because it allows us to approach the process of note-taking and interpreting while contributing ample empirical data.

Chapter 3 sets the theoretical and methodological foundation for the thesis. It defines and operationalises the construct of cognitive load in interpreting and its measurement. Some techniques that could be used to study cognitive processing and cognitive load in CI are also introduced, with special emphasis put on the combination of pen recording and eye tracking.

Chapter 4 reports the data collected in a pilot study of the PhD project. Through a carefully-selected sample of five professional interpreters, this pilot study is able to find some empirical evidence concerning the important variables identified in Chapter 2. Pen recording is proven to be a powerful method to tap into the process of note-taking and interpreting, and the collected data serve as useful indicators of cognitive load. This article paves the way for the next stage of the project where a main study is conducted, combining pen recording and eye tracking and involving a larger group of professional interpreters. Findings of this study are instrumental in determining the hypotheses for the next stage of the study.

Chapter 5 and Chapter 6 report the data collected in the main study of the PhD project. Chapter 5 reports the pen and voice recording data collected from professional interpreters while they perform CI with notes, with special attention paid to Phase I where interpreters listen to the source speech and take notes. Most of the findings in the pilot study have been successfully replicated in the main study. Pedagogical recommendations are provided on the basis of the empirical findings.

Chapter 6 reports the eye tracking data collected from the main study. The focus is on Phase II, in which interpreters read back their notes and produce a target speech. This study is among the first to visualise note-reading, showing that it proceeds in a non-linear manner and requires a high level of cognitive costs. The data provide important indicators of cognitive processing in note-reading and interpreting. A combined analysis of findings from the fourth and fifth article reveals that there is a trade-off between the cognitive costs of Phase I and Phase II in CI.

Chapter 7 summarises the main findings of the research, explains their implications, examines the strengths and limitations of the research, and discusses the possibilities for future work.

### 1.5 References


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting, 14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199–206). Amsterdam: John Benjamins.

Allioni, S. (1989). Towards a grammar of consecutive interpretation. In L. Gran & J. M. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation* (pp. 191–197). Udine: Campanotto.

Barik, H. C. (1973). Simultaneous interpretation: Temporal and quantitative data. *Language and Speech, 16*(3), 237–270. doi:10.1177/002383097301600307

Becker, W. (1972). *Notizentechnik [Note-taking]*. Germersheim: BBK.

Cardoen, H. (2013). The effect of note-taking on target-text fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging research in translation studies: Selected papers of the CETRA Research Summer School 2012* (pp. 1–22). Leuven: CETRA.

Christoffels, I. K. (2004). *Cognitive studies in simultaneous interpreting* (Unpublished doctoral thesis), University of Amsterdam, Amsterdam.

Christoffels, I. K., & De Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition, 7*(3), 227–240. doi:10.1017/S1366728904001609

Christoffels, I. K., & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). New York: Oxford University Press.

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research, 39*(2), 136–144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting, 6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters' notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V. (2010). Consecutive interpreting. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies: Volume 1* (pp. 75–79). Amsterdam: John Benjamins.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (Vol. 7, pp. 227–254). Berlin: Mouton de Gruyter.

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation: Three approaches. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 25–56). London and New Delhi: Thousand Oaks.

Gerver, D. (1974a). The effects of noise on the performance of simultaneous interpreters: Accuracy of performance. *Acta Psychologica, 38*(3), 159–167.

Gerver, D. (1974b). Simultaneous listening and speaking and retention of prose. *Q J Exp Psychol (Hove), 26*(3), 337–341.

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. B. W. Anderson & R. W. Brislin (Eds.), *Translation: Applications and research* (pp. 165–207). New York: Gardner Press.

Gile, D. (1991). Prise de notes et attention en début d'apprentissage de l'interprétation consécutive–une expérience– démonstration de sensibilisation [Note-taking and attention at the beginning of consecutive interpreting learning –an experience–demonstration of awareness]. *Meta*, *36*(2–3), 431–439.

Gile, D. (2001). The role of consecutive in interpreter training: A cognitive view. Retrieved 26 November 2016, from AIIC http://aiic.net/page/377/

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

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. Manchester and Kinderhook: St. Jerome.

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

Gran, L. (1982). *L'annotazione grafica nell'interpretazione consecutiva [Note-taking in consecutive interpreting]*. Trieste: Università degli Studi di Trieste.

Henderson, J. A. (1976). Note-taking for consecutive interpreting. *Babel*, *22*(3), 107-116.

Hill, J. K. (1979). Consecutive interpreting in advanced language work. *Meta*, *24*(4), 442-450. doi:10.7202/003071ar

Ilg, G. (1988). La prise de notes en interprétation consécutive. Une orientation générale [Note-taking in consecutive interpretation: A general approach]. *Parallèles*, *9*, 9–13.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Kohn, K., & Albl-Mikasa, M. (2002). Note-taking in consecutive interpreting: On the reconstruction of an individualised language. *Linguistica Antverpiensia*, *1*, 257-272.

Köpke, B., & Signorelli, T. M. (2012). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism, 16*(2), 183-197. doi:10.1177/1367006911402981

Lambert, S. (1988). Information processing among conference interpreters: A test of the depth-of-processing hypothesis. *Meta, 33*(3), 377-387. doi:10.7202/003380ar

Matyssek, H. (1989). *Handbuch der notizentechnik für dolmetscher [Handbook of note-taking for interpreters]*. Heidelberg: Julius Groos.

Moser-Mercer, B. (1997). Beyond curiosity: Can interpreting research meet the challenge? In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 176–195). London and New Delhi: Thousand Oaks.

O'Donnell, C. R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, vol. II: Cognitive processes and performance* (pp. 42-41–42-49). New York: Wiley-Interscience.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371. doi:10.1007/BF02213420

Paneth, E. (1984). Training in note-taking (For interpreting). In W. Wilss & G. Thome (Eds.), *Die theorie des übersetzens und ihr aufschlußwert für die übersetzungs- und dolmetschdidaktik [The theory of translation and its value for translating-interpreting dichotomy]* (pp. 326–332). Tübingen: Gunter Narr.

Pöchhacker, F. (2004). *Introducing interpreting studies*. London and New York: Routledge.

Rozan, J.-F. o. (1956/2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Cracow: Tertium Society for the Promotion of Language Studies.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting, 13*(2), 176–204. doi:10.1075/intp.13.2.02see

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target, 25*(1), 18–32. doi:10.1075/target.25.1.03see

Seleskovitch, D. (1975). *Langage, langues et mémoire: étude de la prise de notes en interprétation consécutive [Speech, language, and memory: A study of note-taking in consecutive interpreting]*. Paris: Minard Lettres Modernes.

Shlesinger, M. (2000). *Strategic allocation of working memory and other attentional resources in simultaneous interpreting* (Doctoral dissertation), Bar-Ilan University, Israel.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York: Springer.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting*, *8*(2), 129–147. doi:10.1075/intp.8.2.02sza

## An introductory note to Chapter 2

As has been elaborated in Chapter 1, note-taking as a unique characteristic of CI has attracted the interest of researchers for over half a century. The earlier prescriptive literature has established an array of note-taking systems, principles of which are still useful today. The descriptive literature has inspected specific variables concerning note-taking and CI, contributing valuable empirical data for a deeper understanding of the topic. This rich dichotomy of literature contains valuable information on how note-taking research has evolved over the years and what awaits future studies.

Chapter 2 therefore consists of a critical literature review for the thesis. It combs through the prominent and influential studies in the field, with an aim to identify the gaps and limitations in previous research and to inform how investigations should be carried out in this study. It examines several issues that are of central importance to the PhD thesis: the major topics in previous note-taking studies, the research methods that have been used, and the findings, controversies, and limitations of these studies. The chapter also pinpoints the study of cognitive processing and cognitive load as a productive avenue for future research.

# Chapter 2 Note-taking in consecutive interpreting: A review with special focus on Chinese and English literature[1]

**Abstract:** Publications on note-taking in consecutive interpreting are reviewed, with special attention being awarded to literature written in Chinese and English. The review identifies two main streams of note-taking literature, a prescriptive stream and a descriptive stream. Prescriptive publications are concerned with the question "How should notes be taken?" They introduce the established note-taking systems and principles, and discuss how to teach them to students. The second stream, consisting of descriptive studies, tackles the question "How are notes taken?" The studies strive to approach the topic with scientific rigor by collecting data from simulated interpreting practices. Fruitful results have been created, but there are several limitations. The prescriptive stream lacks systematic empirical research to support the proposals. The descriptive stream is mostly product-oriented, lacking process research, and no study has designed true experiments to explain the causal relationships behind the observed phenomena. Cognitive load offers a promising perspective to approach the process of note-taking while contributing ample empirical data. It is therefore worthwhile to investigate cognitive load during note-taking in consecutive interpreting.

**Keywords:** Note-taking, consecutive interpreting, review, cognitive load.

## 2.1 Introduction

Note-taking is a distinctive feature of consecutive interpreting (CI), in particular 'classic' consecutive where systematic note-taking is used (Pöchhacker, 2004, p. 19), and scholars' sustained interest in the subject has generated a considerable volume of literature. This review attempts to explore how note-taking literature has evolved for over half a century, and what awaits future research. It is part of a larger project that looks into the process of CI and note-taking. The aim is to identify the most productive avenue of investigation by combing through the prominent and influential studies in the field. The review is interested in the following questions: (1) what are the major topics in note-taking studies, (2) what research methods have been used, (3) what are the

[1] This chapter is a published journal article: Chen, S. (2016). Note-taking in consecutive interpreting: A review with special focus on Chinese-English literature. *The Journal of Specialised Translation*, 26, 151–171.

findings, controversies, and limitations, and (4) what could be a promising avenue for future research.

The review is written from the perspective of an interpreting researcher and practitioner whose interest lies in interpreting between the Chinese and English language pair, and as a result, the review pays special attention to publications written in Chinese and English[2]. The selection of texts to be reviewed followed a number of successive steps. The inclusion criteria might seem overly restrictive, but they were designed to identify the most relevant and representative literature, not to comprehensively survey all the studies available.

Firstly, the key databases relevant to this review were identified: the Conference Interpreting Research Information Network Bulletin (CIRIN Bulletin), Translation Studies Abstracts Online (TSA Online), and the Chinese Social Sciences Citation Index (CSSCI). The reason why an additional China-based database is necessary for the search of relevant literature is that, according to Zhou, Su, and Leydesdorff (2010, p. 1362), “the Chinese and international communication systems in the social sciences are almost completely uncoupled in terms of the coverage in the databases.”

Secondly, the selected databases were searched for relevant publications. The CIRIN Bulletin did not provide keywords of its collections, so a keyword-based search was not possible. The database was searched for titles that contained any of the following words: ‘note,’ ‘notation’ and ‘note-taking.’ TSA Online and the CSSCI were searched by looking for titles and keywords that contained ‘consecutive interpreting’[3] and any of the above mentioned words.

Thirdly, the references of the retrieved items were scanned for relevant publications. This was repeated until no more relevant publications came to light.

A further step was to finalise the list using the following criteria: (1) The publication addressed note-taking in CI as a subject in its own right[4]. (2) The publication was written in Chinese or English. For publications written in other languages, only those that have been most referenced were included in the review. (3)

[2] It has to be admitted that the author does not have the linguistic prerequisites to read all that she has collected, and that what she has collected represents a comprehensive yet not complete list of the relevant publications.

[3] This is not necessary for the CIRIN Bulletin because the database collects only interpreting studies.

[4] Many authors include note-taking as a part of their discussions on interpreter training or education. These publications are not included in this review.

The study was peer-reviewed (published as a book or book chapter; journal article; in refereed conference proceedings)[5].

The selected publications are grouped according to topic: note-taking systems and principles, didactics, cognitive and linguistic aspects of note-taking, choice of form and language, and the relationship between note-taking and interpreting quality. The boundaries between the groups are by no means clear-cut. The classification has been made on the basis of the most prominent focus of each. Two main streams of literature have been identified: a prescriptive stream and a descriptive stream. Prescriptive publications on note-taking are concerned with the question "How should notes be taken?" Usually starting from the authors' experience in the profession and/or in teaching, this first stream aims at introducing established note-taking systems and principles, and discusses how to teach them. The second stream, consists of descriptive studies, tackles the question "How are notes taken?" It strives to approach the topic with scientific rigor by collecting data from simulated interpreting practices, using either students or professional interpreters as subjects. As will be made clear in the review, there is a shift in note-taking literature from prescriptive to descriptive along this continuum.

## 2.2 Note-taking systems and principles: a prescriptive starting point

Among the first publications on note-taking are a number of books and articles that introduce the well-known note-taking systems and principles. These publications adopt a prescriptive stance, and propose the ways that notes should be taken. More often than not, the prescriptions made are based on the author's experience as professional interpreters and/or teachers. For example, Rozan (2002, p. 11) mentioned in the introduction of his book, that "This system is the product of 10 years as a practising interpreter and 4 years teaching the profession." Little if any empirical data has been collected in this stream of literature. Nevertheless, the contributions are obvious: they have offered important experience and knowledge of the profession, and are therefore fundamental to note-taking research.

[5] A number of relevant studies are in the form of unpublished master's and doctoral theses, but the scope of this article does not allow the inclusion of those studies. Interested readers can find the studies reported in various issues of the CIRIN Bulletin at www.cirinandgile.com.

The earliest note-taking system was proposed by Rozan in 1956. The influence of the work is far-reaching. When it was translated into English and Polish in 2002, the editors commented that "it would be hard to find an interpreter in Western Europe whose note-taking style owes nothing to Rozan" (Rozan, 2002, p. 7). Following Rozan, many books and articles on note-taking systems and principles were published in different languages, each generating a profound influence in its own country and some even reached beyond. Some outstanding examples would be Allioni (1989), Becker (1972), Gillies (2005), Gran (1982), Ilg (1988), Kirchhoff (1979), and Matyssek (1989).

When new systems are introduced, authors usually build on the wealth of the previously-established ones, adapting the existing rules as they see fit. To avoid repetition, this part of the review starts from Rozan's system, uses it as a reference, and discusses some of the best-known principles of note-taking. These principles can be found in most of the existing note-taking systems, and different authors have contributed to them.

The first principle is at the core of almost all note-taking systems: noting the idea and not the word. It has been variously expressed as "comprehension" before note-taking (e.g. Deng, 1991, p. 285; Jia, 1995, pp. 77-78) or "analysis" before note-taking (e.g. Alexieva, 1994, p. 206; Chuang, 2008, p. 95; Han, 2002, pp. 25-26; Mu & Lei, 1998, pp. 82-83). This principle emphasises that what is important in note-taking is the idea or "concept" (Gillies, 2005, p. 53) that lies under the actual words used. When taking notes, interpreters should arrive at the underlying meaning through analysis and comprehension of the source speech.

Rozan's second principle consists of the rules of abbreviation. The most important rule, according to Rozan Rozan (2002, p. 16), is that long words (more than 4 to 5 letters) should not be written in full. It is generally suggested that the first and last letters should be used to abbreviate the word, with the latter written as superscript (Gillies, 2005, p. 130; Matyssek, 1989, p. 115; Rozan, 2002, p. 17; Schweda-Nicholson, 1993, p. 200). Using the first letters to abbreviate is also recommended (Becker 1972: 30). Other rules of abbreviation include: using abbreviations to indicate gender, tense and register (Rozan, 2002, pp. 17-18); borrowing commonly known abbreviations from daily life (Matyssek, 1989, p. 113; Wu, 2008, p. 8); using international suffixes such as "-tion" (Gillies, 2005, p. 130; Matyssek, 1989, p. 117); and using phonetic spelling and misspelling (Gillies, 2005, p. 131; Han, 2002, p. 26). It is common to put abbreviations at a prominent place when discussing note-taking in CI between European languages, but the case with Chinese is different. Some of the rules which are largely based on European languages are difficult for native speakers of Chinese, and some rules are even rendered useless because of the differences in languages (M. Liu, 2008, p. 65f). That being said, part of the rules are still applicable, especially when the task is interpreting from Chinese into English. No matter how abbreviations are used, they have to meet certain conditions: they should be unambiguous (Henderson, 1976, p. 110; Matyssek, 1989, p. 115), easy to write (Alexieva, 1994, p. 204), and should not sacrifice accuracy (Schweda-Nicholson, 1990, p. 140).

The third principle concerns the noting of links. Links are believed to be indispensable in note-taking (Matyssek, 1989, p. 53; Wu, 2008, p. 17) because “an idea can be distorted completely if its relation to the previous idea is not clearly indicated” (Rozan, 2002, p. 18). Many authors (e.g. Gillies, 2005; Matyssek, 1989; Wu, 2008) have identified the main types of linking words and expressions, including additive, adversative, and causal (cause, purpose and consequence) links, and recommended the use of only one abbreviation, short word, or symbol to represent the whole family. Gillies (2005, pp. 147, 149) also points out the importance of adding implicit links and dropping link words that are not links.

Rozan’s fourth and fifth principles refer to the noting of negation and emphasis. Negation is usually achieved by crossing out, and emphasis by underlining (Gillies, 2005, p. 106; Matyssek, 1989, pp. 107-110; Rozan, 2002, p. 19; Schweda-Nicholson, 1993, pp. 201-202). Emphasis could also be achieved by shifting, i.e. moving notes further to the left or right on the notepad (Gillies, 2005, p. 83).

The last two on Rozan’s list are the principles of verticality and shift, the “backbone” of his system (Rozan, 2002, p. 20). These two are fundamentally principles on the layout of the notes, and have been given different names by other authors, such as the use of space (M. Liu, 2008, p. 52) and diagonal layout (Jones, 1998, p. 44; Özben, 1993, p. 42). According to the principles, notes should be structured in a “vertical, indented and terraced way” (Kohn & Albl-Mikasa, 2002, p. 262), so that the units of meaning are easy to identify when reading back notes. A mind-mapping note-taking technique which starts from the centre of the page is also proposed by Torres Díaz (1997).

Another important part in any existing note-taking system is the use of symbols. Symbols are used because they are easy to write and read, and represent concepts not words, thus avoiding source language influence (Gillies, 2005, p. 99). Distributed towards the two ends of the minimalist-maximalist continuum of symbols are Rozan and Matyssek. The former recommended a total of 20 symbols, of which "only 10 were indispensable" (Rozan, 2002, p. 25), while the latter used a whole book volume to introduce a detailed code of drawings and symbols. Although Matyssek's system was sometimes criticised as running the risk of becoming an "interpreter's shorthand" (Ilg & Lambert, 1996, p. 72), he emphasised that the symbols were suggestions rather than obligatory requirements (Matyssek, 1989, p. 233). Moreover, an in-depth analysis into the two systems by Ahrens indicated that they "do not differ at all as far as the basic principles of note-taking are concerned" (Ahrens, 2005, p. 13). Other authors are more or less distributed along the continuum, suggesting more symbols than Rozan, but rejecting the idea of using a symbol-based note-taking system. Generally speaking, symbols are believed to be very helpful when they are simple, unambiguous, and fully mastered by the interpreter. It is also pointed out by many authors that it should be possible to combine symbols to create new symbols (Allioni, 1989; Gillies, 2005; Matyssek, 1989; Wu, 2008).

So far, it would seem that the principles of note-taking are well-developed, and once the students are made aware of them and practice accordingly, note-taking should not be a problem at all. However, when it comes to teaching and learning these principles, both the students and teachers find it challenging.

### 2.3 Note-taking didactics: the beginning of a shift from prescriptive to descriptive

With effective note-taking principles having been worked out and applied by eminent professionals, two problems now arise: the first is whether these principles and systems can be taught to students; and if so, the second is how note-taking can be taught systematically.

The individuality of any note-taking system is emphasised by all who have written on the topic. This is why some authors do not believe in the systematic teaching of note-taking. The case in France is typical of this attitude. As Ilg and Lambert pointed out, "The École Supérieure d'Interprètes et de Traducteurs (ESIT, Paris) never thought much of note-taking as an underpinning of CI", and the publications "were sketchy as far as the techniques of CI are concerned" (Ilg & Lambert, 1996, p. 71). Thiéry (1981) was an example of this sceptical attitude towards teaching note-taking systematically. He argued that instructions on note-taking should be limited only to essentials, and that systematic note-taking as a creative and individual activity, could not be taught.

Nevertheless, many authors believe note-taking should be taught systematically to students, and they have made great efforts to operationalise their didactic proposals. The discussions target three different student groups: post-graduate level interpreting students, undergraduate language students, and community interpreters.

The discussions begin with note-taking training for potential candidates of the profession, usually at post-graduate levels. In fact most of the above-mentioned literature on note-taking systems and principles fall into this category. Apart from the publications that focus exclusively on note-taking and treating it as a subject in its own right, there are also a large quantity of literature that has addressed note-taking as part of the discussions on interpreter training. Those discussions however, go beyond the scope of this article. Interested readers are referred to such authors as Bowen and Bowen (1980), Ilg and Lambert (1996), Jones (1998), Kunihiro *et al.* (1969), Seleskovitch and Lederer (1989/1995), Schweda-Nicholson (1985), van Hoof (1962) and Zhong (1999).

With interpreting being taught to more and more undergraduate language students as a language reinforcement activity, many teachers have detected the differences in this new group (e.g. no aptitude testing before entering the classes and great student numbers), and discussed how to make adaptations accordingly (e.g. Weiwei Dai & Xiang, 2008; Henderson, 1976; Her, 2001; Paneth, 1984).

Teaching note-taking to community interpreters is uniquely addressed by Schweda-Nicholson (1990), who is interested in those natural bilinguals without much specialised training. The goal was to enable community interpreters to benefit from note-taking by teaching them the basic techniques.

Differences in the type of students lead to differences in the teaching objectives and choice of materials (e.g. Henderson, 1976; Her, 2001), but the fundamental training rationales are quite similar. Teachers are well aware that note-taking could take away attentional resources from other activities in the interpreting process and cause problems. They usually advise the students against taking notes in the beginning stage of training. Instead, much attention is devoted to a series of other exercises such as speech analysis, summarising exercises, and memory training. Actual note-taking is only introduced after a period of those trainings, and students begin practice with easy materials so that they are not overwhelmed by the multi-tasking. Gillies (2005) even suggested practising with written materials (transcripts of speeches) rather than spoken ones in the initial stage.

However, despite the awareness of the difficulties and the precautions taken, both the teachers and students still find it challenging to teach/learn note-taking. Studies that describe the difficulties met by students in classes represent the beginning of a shift from prescriptive to descriptive stream in note-taking research.

Gile (1991) divided 14 students evenly into two groups for CI exercises containing proper nouns. One group was instructed to take notes and the other was refrained from doing so except for names and figures. He found that the note-taking group heard the names worse, and explained that it was because note-taking diverted attention from listening and led to a degradation of listening quality.

A longitudinal study by Alexieva showed that the instruction in note-taking systems and principles “brings about a trough in students’ performance, which remains consistently low for a comparatively long period” (1994, p. 200). The same phenomenon was found by Her (2001, p. 62). Alexieva (Alexieva, 1994, p. 200) inferred that at this stage, note-taking learning was characterised by “a weaker memory operational capacity,” because most of the students’ energy was spent on deciding what symbols to use, recalling the symbols, and deciding what to put in notes and what to put in memory.

To see how difficulties were perceived by students in note-taking, Xu and Chai (2008) used stimulated recall and post-task interviews to investigate the issue. The major difficulties reported include: insufficient memory, inadequate recall when using notes as cues, improper form of notes, and overdependence on notes without proper processing of source information.

Chmiel was interested in the effectiveness of note-taking teaching, and put students to a test after a note-taking course. The overall results were “less encouraging than expected” (Chmiel, 2010, p. 248), with the techniques taught in the course being applied in only 57% of the cases. She also found that layout and visualisation techniques were more readily transferable than symbols to students’ individual note-taking systems.

Also interested in evaluating learning outcomes, Orlando (2010) made a technological contribution to the didactic advancement. He pointed out the deficit in the product-oriented evaluation method, and suggested the application of digital pens, a technology that allows easy recording of the process of note-taking. The questionnaire results he collected from students showed encouraging potentials of the technology in classes.

The studies reviewed in this section represent an early descriptive stance taken by researchers. Instead of simply prescribing how notes should be taken, the authors set out to observe and describe how notes are actually taken by students. This shift from prescriptive to descriptive research is strengthened by scholars who approach the topic of note-taking from linguistic and cognitive perspectives.

## 2.4 Cognitive and linguistic aspects of note-taking: a theoretical drive

Investigations on the cognitive and linguistic aspects of note-taking are mainly motivated by an attempt to theorise note-taking and CI. The two pioneering authors and their investigations (Kirchhoff, 1979; Seleskovitch, 1975) were certainly ahead of their time.

Seleskovitch (1975) set out to develop a theory to systematise the ESIT's training methods. She conducted an experiment in which she collected and analysed the notes taken by 12 professional interpreters. She found that the notes included few of the words in the source speech and many outside the speech, that the renditions expressed much more than the notes, and that some items appeared in different forms. Based on the findings, she inferred the formal independence of the source speech, notes, and target speech, pointing to an intermediate stage of "deverbalisation." Her cognitive model of interpreting assigned linguistic and cognitive processing to different kinds of memory, and pointed out that notes functioned as minimal memory triggers, rather than "an exhaustive code" (Setton, 2002, p. 119).

Standing in contrast to Seleskovitch's deverbalised view towards note-taking, Kirchhoff (1979) was concerned about the linguistic surface structures of the notes. She saw notes as a kind of physical storage as opposed to the cognitive storage of memory. Note-taking was believed to be a primarily linguistic process, based on the microstructures of the source text. Her view of notes as a type of language was supported and followed by Albl-Mikasa, who looked into the language and discourse dimensions of consecutive notes (Kohn & Albl-Mikasa, 2002), the reduction and expansion processes in note-taking and note-reading (Albl-Mikasa, 2006) and how interpreters worked closely along micro-propositional lines when processing the source, notation and target texts (Albl-Mikasa, 2008). The authors believe that, although the fundamental principle of note-taking is noting the idea and not the word, note-taking usually operates on a micro-level that stays close to the source text.

Despite the difference in stress (in sense or in linguistic surface structure), the scholars have consistently pointed out a concurrent storage of information in memory and in notes, as well as a competition for cognitive resources between note-taking and other activities in CI, an issue at the core of Gile’s (2009) Effort Models of interpreting.

Giles’s Effort Model of consecutive interpreting conceptualises the interpreting process in two phases: a comprehension (or listening and note-taking) phase, and a speech production (or reformulation) phase. The model assumes four processing capacity demands, or “Efforts” (2009, p. 160) in the first phase, each relating to a specific activity in the process: Listening and analysis, Note-taking, Short-term memory operations, and Coordination. In the second phase there are three Efforts: Remembering, Note-reading, and Production. The Efforts are competing and processing capacity is limited. In order for interpreting to proceed smoothly, the total processing capacity demands should not exceed the available capacity, and each Effort should not exceed the available capacity for each activity. Gile believes note-taking is critical for CI in terms of cognitive capacity, and the key lies in “how to reduce processing capacity and time requirements of note-taking while maintaining the efficiency of notes as memory reinforcers” (2009, p. 178).

Gile’s model, though originally developed to inform teaching, is found useful by many scholars in academic research. It is mentioned in various explorations on the prominent features of note-taking.

## 2.5 Exploring the key note-taking features: descriptive studies on notes and quality

Unlike the early empirical investigations which have a general interest in what real notes look like, and set out to discover some overall trends, studies reviewed in this section have more specific targets. They usually focus on certain note-taking features, and conduct experiments to closely investigate the features of interest. They have contributed the largest quantity of empirical data on the topic to date. The most important variables explored are: the choice of form, the choice of language, and the relationship between note-taking and interpreting quality.

### 2.5.1 The choice of form and language in note-taking

The choice of form in note-taking refers to the choice between language and symbol, and the choice between abbreviation and full word; while the choice of language refers to the choice between source and target language, and the choice between A and B language[6].

A rare and detailed video documentation of note-taking was compiled by Andres (2002). She recorded the note-taking processes of 14 professionals and 14 students interpreting from French to German. The notes of the two groups were compared, and Andres found that despite a source language preference in both groups, the professional group wrote more target language units than the student group. She also used the time-coded videos to study time lags in note-taking. According to her findings, the time lag between listening and note-taking was three to six seconds for professionals, while reaching as much as ten seconds for students. Her findings provided abundant evidence of processing overload in students during the first phase of interpreting.

The most comprehensive series of studies to date on note-taking features were conducted by Dam and her colleagues (Dam, 2004a, 2004b, 2007; Dam, Engberg, & Schjoldager, 2005). Dam's study (2004b) with notes taken by four students shows that the choice of language in note-taking is largely governed by the A/B language status, rather than the source/target one, with all participants preferring A language regardless of the direction of interpreting. Her study with five professionals revealed that the participants' preferences for the form of note-taking were: symbols (41% of all note units), followed by full words (35%) and abbreviations (25%) (Dam, 2004a, p. 254). Again, all participants showed a clear preference for target language, their A language. She also found that more notes were taken in the source language when the source text was more difficult.

Dam's studies were based on CI between Danish and Spanish, and that raises questions about the generalisability of her results to other language pairs. Following Dam, other scholars have experimented with different language pairs. Some representative examples are: Lung (2003), Dai and Xu (2007), Liu (2010), and Wang, Zhou, and Wang (2010) with Chinese and English; Lim (2006) with Korean and English; Szabó (2006) with Hungarian and English; and González (2012) with Spanish and English.

[6] In this article, A language refers to the native language while B language refers to the foreign language.

Lung (2003) studied the notes of 21 students interpreting from English to Chinese, and found that the students made little use of either abbreviations or symbols, and that the notes consisted mainly of source and B language. Dai and Xu (2007) looked at the notes taken by 12 students interpreting from Chinese to English, and found that the notes were source and A language dominated. The 120 students in Liu's (2010) experiment on the whole showed a preference for language over symbol, and full word over abbreviation. Wang *et al.* (2010) experimented with 12 students, and the notes were predominantly source language with few symbols used, and abbreviations were used more than full words. Szabó (2006) looked at the notes taken by eight professionals interpreting between Hungarian and English, and discovered that her subjects showed a clear preference for English, their B language, regardless of the direction of interpreting. The results suggested that the language combination itself played an important role in the choice of language. Abuín González (2012) compared the notes taken by three groups of subjects with varying levels of experience (beginner students, advanced students and interpreters) when interpreting from English to Spanish. The results showed a shift in language preference from source to target with an increasing level of expertise.

The details of the studies are summarised in Table 2.1. It is easy to see how they vary greatly in terms of the design (e.g. type of participants, language pair, interpreting direction). Moreover, many studies did not specify the details of the tasks used in the experiment, making it even more difficult to compare the results.

Table 2.1 A summary of studies on key note-taking features

| Studies | Participants: Number | Participants: Prof/Stud | Tasks: Language pair | Tasks: Direction | Tasks: Length | Tasks: Duration | Tasks: Speed | Results: Form | Results: Language |
|---|---|---|---|---|---|---|---|---|---|
| Andres 2002 | 28: 14 | Prof | German A French B | B to A | NS | 6'27" | 167.9 syllabus per minute | Language>Symbol | Source>Target B>A |
| | 14 | Stud | | | | | | | |
| Lung 2003 | 21 | Stud | Chinese A English B | B to A | NS | | | Language>Symbol Full word>Abbreviation | Source>Target B>A |
| Dam 2004a | 4: 3 | Stud | Danish A Spanish B | Both directions | NS | | | Language>Symbol | A>B |
| | 1 | Stud | Spanish A Danish B | | | | | | |
| Dam 2004b | 5 | Prof | Danish A Spanish B | B to A | 1081 words | 7'30" | NS | Language>Symbol Full word>Abbreviation | Target>Source A>B |
| Lim 2006 | 40 | Stud | Korean A English B | A to B | NS | | | NS | Source>Target A>B |
| Szabó 2006 | 8 | Prof | Hungarian A English B | Both directions | NS | | | NS | B>A |
| Dai & Xu 2007 | 12 | Stud | Chinese A English B | A to B | 529 characters | NS | | Language>Symbol Abbreviation>Full word | Source>Target A>B |
| Liu 2010 | 120 | Stud | Chinese A English B | A to B | 518 characters | NS | 200 characters per minute | Language>Symbol Full word>Abbreviation | |
| Wang *et al.* 2010 | 12 | Stud | Chinese A English B | Both directions | 135 words / 163 characters | NS | | Language>Symbol Abbreviation>Full word | Source>Target |
| Abuín González 2012 | 30: 10 | Prof | Spanish A English B | B to A | 711 words | 6'22" | 125 words per minute | NS | Prof: A>B Target>Source |
| | 20 | Stud | | | | | | | Stud: B>A Source>Target |

Notes: Prof=Professionals, Stud=Students, A=A language, B=B language, NS=Not Specified, Source=Source language, Target=Target language

Neverthless the author believes it could be beneficial to try and compare the findings on each note-taking feature (i.e. the choice of form and the choice of language), and see if some general trends could be detected. Results on the choice of form, as presented in Table 2.2, point to a dominance of language over symbol, and a slight tendency to use more full words than abbreviations. Results on the choice of language, however, yileds much more inconsistent findings.

Table 2.2 Findings on the choice of form in note-taking

| Studies | Lung 2003 | Dam 2004a | Dam 2004b | Dai & Xu 2007 | Liu 2010 | Wang *et al.* 2010 |
|---|---|---|---|---|---|---|
| Results | Language> Symbol | Language> Symbol | Language> Symbol | Language> Symbol | Language> Symbol | Language> Symbol |
| | Full word> Abbreviation | | Full word> Abbreviation | Abbreviation> Full word | Full word> Abbreviation | Abbreviation> Full word |

To reveal the trends in the choice of language in a clearer way, Table 2.3 organises the studies according to the type of participants and interpreting direction. While the language choices of professionals still appear greatly varied, the choices made by students are obviously source-language dominated. This could be explained using Gile's Effort Model. The skills of students are not fully developed, so note-taking consumes a considerable amount of processing capacity, leaving less available for producing target-language equivalents during the note-taking phase. As a result, students opt for source language notes to avoid saturation during the first phase. In the second phase, since it is self-paced, the students have extra time and processing capacity to deal with the translation.

Table 2.3 Findings on the choice of language in note-taking

| **Participants: Students** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Andres 2002** | **Lung 2003** | **Dam 2004a** | **Lim 2006** | **Dai & Xu 2007** | **Liu 2010** | **Wang *et al.*2010** | **Abuín González 2012** |
| **Language pair** | | German, French | Chinese, English | Danish, Spanish | Korean, English | Chinese, English | Chinese, English | Chinese, English | Spanish, English |
| **Direction** | **A to B** | | | Source> Target A>B | Source> Target A>B | Source> Target A>B | Source> Target A>B | Source> Target A>B | |
| | **B to A** | Source> Target B>A | Source> Target B>A | Target> Source A>B | | | | Source> Target B>A | Source> Target B>A |

| **Participants: Professionals** | | | | | |
|---|---|---|---|---|---|
| | | **Andres 2002** | **Dam 2004b** | **Szabó 2006** | **Abuín González 2012** |
| **Language pair** | | German, French | Danish, Spanish | Hungarian, English | Spanish, English |
| **Direction** | **A to B** | | | Target> Source B>A | |
| | **B to A** | Source> Target B>A | Target> Source A>B | Source> Target B>A | Target> Source A>B |

Notes: A=A language, B=B language, Source=Source language, Target=Target language

What is also made clear in Table 2.3 is that, despite the efforts to describe how notes are acually taken, there is a lack of research done with professional interpreters. However, in order to observe know how notes are acually taken in consecutive interpreting, it is necessary to observe the behaviours of practicing interpreters, rather than students who have not fully mastered the technique. The same weakness could be detected in studies on the relationship between note-taking and interpreting quality.

### 2.5.2 The relationship between note-taking and interpreting quality

Having observed the greatly varied features of note-taking, some researchers begin to empirically investigate the relationship between these features and the quality of interpreting performance. Most of the studies use student interpreters as participants, because quality is an issue at the core of the teaching of interpreting.

Dam *et al.* (2005) generated hypotheses about features of efficiency and non-efficiency in notes, based on their proposal to judge the accuracy of the target text through analysing the semantic network. The hypotheses were later tested by Dam (2007) with notes taken by five professionals interpreting from Spanish to Danish. She found evidence for two of the hypotheses: "the more notes, the better the target text – and vice versa," and "the more abbreviations/the fewer full words, the better the target text – and vice versa," but the data failed to support the third hypothesis: "the more notes in the source language/the fewer in the target language, the better the target text" (2007, p. 194).

Experimenting on the language pair of Chinese and English, Her (2001) analysed the notes taken by undergraduate students interpreting between Chinese and English. She found that there was a general positive relationship between the quality of notes and the quality of interpreting, although good notes did not necessarily yield good performance. Dai and Xu (2007) were unable to find evidence for Dam's (2007) hypotheses. Their data showed that an increase in the quantity of notes did not necessarily mean better target text. Similar conclusions were reached by Liu (2010), who found no significant difference in the quantity or language of the notes taken by high- and low-score groups. But he was able to observe that the high-score group used more symbols than the low-score group. Wang *et al.* (2010) also found no significant relationship between interpreting quality and the quantity, form or language of note-taking. The fact that Dam's findings were not replicated in the above studies might partly be explained by the participants used: Dam used professional interpreters, while the others used students.

Also using students as participants, a study by Cardoen (2013) found relationships that were opposite to Dam's findings. Three participants interpreted from Spanish to Dutch, and Cardoen found that fluent chunks contained fewer notes, more full words and fewer abbreviations when compared with disfluent chunks.

Studies reviewed in this section are summarised in Table 2.4. They have used different types of participants and tasks, and they do not always specify the details of their design. Based on what has been collected so far, it would seem that the interactions between note-taking and interpreting quality are more complex than researchers have imagined.

Table 2.4 A summary of studies on the relationship between note-taking and interpreting quality

| Studies | Participants | | Tasks | | | | | | Results |
|---|---|---|---|---|---|---|---|---|---|
| | Number | Prof/ Stud | Language pair | Direction | Length | Segments | Duration | Speed | |
| Her 2001 | 27 | Stud | Chinese A English B | Both directions | NS | 5 | 5' | NS | Better notes, better quality |
| | | | | | | 2 | 3'15" | 80 cpm | |
| Dam et al. 2005 | 1 | Prof | Danish A Spanish B | B to A | | NS | | | More notes, more abbreviations/fewer full words, more source language/fewer target language, better quality |
| Dam 2007 | 5 | Prof | Danish A Spanish B | B to A | 1092 words | | NS | | More notes, more abbreviations/fewer full words, better quality |
| Dai & Xu 2007 | 12 | Stud | Chinese A English B | A to B | 529 characters | 5 | NS | | No significant relationship |
| Liu 2010 | 62 | Stud | Chinese A English B | A to B | 518 characters | NS | | 200 cpm | More symbols, better quality |
| Wang et al. 2010 | 12 | Stud | Chinese A English B | Both directions | 135 words | 5 | NS | | No significant relationship |
| | | | | | 163 characters | 5 | | | |
| Cardoen 2013 | 3 | Stud | Dutch A Spanish B | B to A | | NS | | 96 wpm | Fewer words, more full words/fewer abbreviations, better quality |

Notes: Prof=Professionals, Stud=Students, A=A language, B=B language, NS=Not Specified, cpm=characters per minute, wpm=words per minute.

The empirical studies reviewed in this section vary greatly in terms of their design (as made evident in Table 2.1 and Table 2.4), and are therefore difficult to compare. Although some general trends can be found, such as a source language dominance in the notes taken by students, and more target language in professional interpreters' notes compared with students, there are also vast inconsistencies. These inconsistencies are a great place to start with for future studies.

## 2.6 Limitations of previous studies

There is no doubt that fruitful results have been created during the past decades, but it is necessary to point out the limitations in order to inform future research endeavours.

In the prescriptive stream, a common limitation is a lack of systematic and rigorous empirical research to support the proposals. It is therefore gratifying to see a shift from prescriptive to descriptive research, with an increase in the quantity of empirical studies. Also, a variety of research methods have been used, such as simulation, case study, questionnaire survey, stimulated recall, and interview. However, a few limitations still exist. First, most of the descriptive studies are product-oriented, but product analysis only allows speculations about the underlying processes based on data collected afterwards. Besides due to the highly individualised nature of interpreting notes, it is often difficult to observe any uniformity in their surface structures. Second, most of the studies use students instead of professional interpreters as participants, and data is collected under simulated rather than real life contexts. But in order to get a better picture of how notes are taken, it is necessary to observe the behaviours of professionals in field interpreting. Third, no study has pushed the shift forward to an explanatory stream. The researchers usually stop at describing what notes look like, but no one has designed true experiments to explain the causal relationships behind the phenomena observed.

In order to initiate a shift from descriptive to explanatory research, an overarching framework is needed to cohesively pull together all the efforts in note-taking studies. It is the belief of the author that a cognitive load perspective towards note-taking has great potentials in that regard.

## 2.7 Cognitive load: a promising avenue for investigation

Interpreting is deemed a cognitively demanding task by different scholars, many of whom have pioneered the investigation of cognitive load in simultaneous interpreting (e.g. Gile, 2008; Hyönä, Tommola, & Alaja, 1995; Seeber, 2011, 2013; Seeber & Kerzel, 2012; Tommola & Hyönä, 1990). Compared to that, research on cognitive load in CI and note-taking seems to be scarce. However, as Gile (2009, p. 178) points out, "note-taking is an area in which the concept of processing capacity can be useful."

Cognitive load is defined by Seeber (2013, p. 19) as "the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system." Starting from a cognitive load perspective, all discussions on note-taking boil down to one fundamental question: how to reduce the cognitive load of note-taking while maintaining the efficiency of notes.

If cognitive load can be measured while interpreters take notes and interpret, some fundamental principles underlying the note-taking choices might be unveiled. For example, it is possible that no matter what choices an interpreter makes (e.g. writing notes in the source or target language), the result is always a lower level of cognitive load for that particular interpreter in that particular task. That is to say, the differences observed in the note-taking behaviours in previous studies might not be controversies, but rather converging evidence in proving that interpreters make choices according to their own situations to reduce cognitive load.

However, measuring cognitive load is no easy task. The construct is generally believed to be multi-dimensional and therefore difficult to measure. Scholars working on the Cognitive Load Theory (e.g. Paas, Tuovinen, Tabbers, & Van Gerven, 2003;

Paas & Van Merriënboer, 1994) specified two dimensions of cognitive load: a causal dimension reflecting the factors that affect cognitive load, and an assessment dimension corresponding to factors that are affected by cognitive load. The assessment factors, including mental load, mental effort, and performance, are indicative of cognitive load, and are therefore used for its measurement. A detailed discussion into the assessment factors and the related measures goes beyond the scope of this article. Interested readers are referred to such works as Paas *et al.* (2003) and Plass, Moreno, and Brünken (2010) for a starting point.

The measurement of cognitive load is not new to the field of interpreting. Many of the studies are overviewed in Seeber (2013). The pioneering studies have laid the groundwork by reviewing important theories, building useful models, discussing methods of measurement, and providing empirical findings. Although the studies have only investigated simultaneous interpreting, much of what has been discussed is also meaningful for CI and note-taking. Hopefully, note-taking research would be able to build on the wealth of those studies and studies in such fields as Cognitive Load Theory, to overcome the limitations faced by previous studies, and to move forward to an explanatory stream of note-taking research.

## 2.8 Bibliography


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting*, *14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Ahrens, B. (2005). Rozan and Matyssek: Are they really that different? A comparative synopsis of two classic note-taking schools. *Forum, 3*(2), 1–15.

Albl-Mikasa, M. (2006). Reduction and expansion in notation texts. In C. Heine, K. Schubert, & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp. 195–214). Tübingen: Narr.

Albl-Mikasa, M. (2008). (Non-) sense in note-taking for consecutive interpreting. *Interpreting, 10*(2), 197–231. doi:10.1075/intp.10.2.03alb

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199–206). Amsterdam: John Benjamins.

Allioni, S. (1989). Towards a grammar of consecutive interpretation. In L. Gran & J. M. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation* (pp. 191–197). Udine: Campanotto.

Andres, D. (2002). *Konsekutivdolmetschen und notation [Consecutive interpreting and note-taking]*. Frankfurt: Peter Lang.

Becker, W. (1972). *Notizentechnik [Note-taking]*. Germersheim: BBK.

Bowen, D., & Bowen, M. (1980). *Steps to consecutive interpretation*. Washington, DC: Pen and Booth.

Cardoen, H. (2013). The Effect of Note-taking on Target-text Fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging Research in Translation Studies: Selected Papers of the CETRA Research Summer School 2012* (pp. 1-22). Leuven: CETRA.

Chmiel, A. (2010). How effective is teaching note-taking to trainee interpreters? *The Interpreter and Translator Trainer*, *4*(2), 233–250.

Chuang, L.-l. (2008). Note-taking know-how: A processing perspective on consecutive interpreting. *Spectrum: Studies in Language, Literature, Translation, and Interpretation*, *2*, 93–101.

Dai, W., & Xiang, C. (2008). 口译教学中记忆、演讲和笔记能力的培养 'The training of memory, speaking and note-taking ability in interpreting teaching'. *Journal of Southwest University for Nationalities, S3*, 101-102.

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research, 39*(2), 136–144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting*, *6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters’ notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected* *contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (Vol. 7, pp. 227–254). Berlin: Mouton de Gruyter.

Deng, Y. (1991). Consecutive notes: An artist's tool. In C.-c. Liu (Ed.), *Collected essays on translation* (pp. 284–292). Hong Kong: Commercial Press.

Gile, D. (1991). Prise de notes et attention en début d'apprentissage de l'interprétation consécutive–une expérience– démonstration de sensibilisation [Note-taking and attention at the beginning of consecutive interpreting learning –an experience– demonstration of awareness]. *Meta, 36*(2–3), 431–439.

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum, 6*(2), 59–77.

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

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. Manchester and Kinderhook: St. Jerome.

Gran, L. (1982). *L'annotazione grafica nell'interpretazione consecutiva [Note-taking in consecutive interpreting]*. Trieste: Università degli Studi di Trieste.

Han, Z. (2002). 笔记在口译中的作用 [The function of note-taking in interpreting]. *Chinese Science and Technology Translators Journal, 15*(2), 25-26, 33.

Henderson, J. A. (1976). Note-taking for consecutive interpreting. *Babel, 22*(3), 107-116.

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation, 6*, 53-77.

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

Ilg, G. (1988). La prise de notes en interprétation consécutive. Une orientation générale [Note-taking in consecutive interpretation: A general approach]. *Parallèles, 9*, 9–13.

Ilg, G., & Lambert, S. (1996). Teaching consecutive interpreting. *Interpreting, 1*(1), 69-99.

Jia, Y. (1995). 口译笔记技巧之我见 [My understanding of note-taking skills in interpreting]. *Foreign Language Education, 1*, 77-80.

Jones, R. (1998). *Conference interpreting explained*. Manchester: St. Jerome.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Kohn, K., & Albl-Mikasa, M. (2002). Note-taking in consecutive interpreting: On the reconstruction of an individualised language. *Linguistica Antverpiensia, 1*, 257-272.

Kunihiro, M., Nishiyama, S., & Kanayama, N. (1969). *Tsuuyaku: Eikaiwakara doujitsuuyakumade [Interpreting: from English conversation to simultaneous interpreting]*. Tokyo: Nihonhousoshuppankyoukai.

Lim, H.-O. (2006). A post-mortem of note-taking. *Forum, 4*(2), 89–112.

Liu, J. (2010). 英语专业本科学生汉英交传笔记特征——一项基于学生交传笔记的实证研究 [Note-taking characteristics of English majored undergraduates in Chinese-English consecutive interpreting: an empirical study based on students' consecutive interpreting notes]. *Foreign Language World, 2*, 47–53.

Liu, M. (2008). *逐步口譯與筆記 [Consecutive interpretation and note-taking]*. Taipei: Bookman.

Lung, R. (2003). Taking "notes" seriously in the interpretation classroom. In Á. Collados Aís, M. M. Fernández Sánchez, & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación [The evaluation of quality in interpretation: Research]* (pp. 199–205). Granada: Comares.

Matyssek, H. (1989). *Handbuch der notizentechnik für dolmetscher [Handbook of note-taking for interpreters]*. Heidelberg: Julius Groos.

Mu, D., & Lei, R. (1998). 试论口译笔记训练中的理解和记忆 [Comprehension and memory in note-taking training]. *Foreign Language Education, 3*, 82–84.

Orlando, M. (2010). Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter, 15*, 71–86.

Özben, R. T. (1993). Considerations on the note-taking process during consecutive interpretation from Italian into Turkish. *The Interpreters' Newsletter, 5*, 42–48.

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. doi:10.1207/S15326985EP3801_8

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371. doi:10.1007/BF02213420

Paneth, E. (1984). Training in note-taking (For interpreting). In W. Wilss & G. Thome (Eds.), *Die theorie des übersetzens und ihr aufschlußwert für die übersetzungs- und dolmetschdidaktik [The theory of translation and its value for translating-interpreting dichotomy]* (pp. 326–332). Tübingen: Gunter Narr.

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive Load Theory*. New York: Cambridge University Press.

Pöchhacker, F. (2004). *Introducing interpreting studies*. London and New York: Routledge.

Rozan, J.-F. o. (2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Cracow: Tertium Society for the Promotion of Language Studies.

Schweda-Nicholson, N. (1985). Consecutive interpretation training: Videotapes in the classroom. *Meta*, *30*(2), 148. doi:10.7202/003731ar

Schweda-Nicholson, N. (1990). Consecutive note-taking for community interpretation. In D. Bowen & M. Bowen (Eds.), *Interpreting: Yesterday, today, and tomorrow* (pp. 136–145). Amsterdam and Philadelphia: John Benjamins.

Schweda-Nicholson, N. (1993). An introduction to basic note-taking skills for consecutive interpretation. In E. Losa (Ed.), *Keystones of communication: Proceedings of the 34th Annual Conference of the American Translators Association* (pp. 197–204). Medford: Learned Information.

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting*, *13*(2), 176–204. doi:10.1075/intp.13.2.02see

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target*, *25*(1), 18–32. doi:10.1075/target.25.1.03see

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism, 16*(2), 228–242. doi:10.1177/1367006911402982

Seleskovitch, D. (1975). *Langage, langues et mémoire: étude de la prise de notes en interprétation consécutive [Speech, language, and memory: A study of note-taking in consecutive interpreting]*. Paris: Minard Lettres Modernes.

Seleskovitch, D., & Lederer, M. (1989/1995). *A systematic approach to teaching interpretation* (J. Harmer, Trans.). Silver Spring, MD: Registry of Interpreters for the Deaf.

Setton, R. (2002). Revisiting the classics. Seleskovitch: A radical pragmatist before her time. *The Translator, 8*(1), 117–124.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting, 8*(2), 129–147. doi:10.1075/intp.8.2.02sza

Thiéry, C. (1981). L'enseignement de la prise de notes en interprétation consecutive: un faux problème? [The teaching of note-taking in consecutive interpretation: A false problem?]. In J. Delisle (Ed.), *L'enseignement de l'interprétation et de la traduction: de la théorie a la pedagogie* (pp. 99–112). Ottawa: University of Ottawa Press.

Tommola, J., & Hyönä, Jukka. (1990). Mental load in listening, speech shadowing and simultaneous interpreting: A pupillometric study. In J. Tommola (Ed.), *Foreign language comprehension and production* (pp. 179–188). Turku: AfinLA.

Torres Díaz, M. G. (1997). Why consecutive note-taking is not tantamount to shorthand writing. In K. Klaudy & J. Kohn (Eds.), *Transferre necesse est. Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September 1996, Budapest, Hungary* (pp. 213–216). Budapest: Scholastica.

van Hoof, H. (1962). *Théorie et pratique de l'interprétation, avec application particulière à l'anglais et au français [Theory and practice of interpreting, with particular application to English and French]*. München: Max Hueber.

Wang, W., Zhou, D., & Wang, L. (2010). 口译笔记特征与口译产出质量实证研究 [An empirical study of note-taking characteristics and output quality in interpreting]. *Foreign Language World, 4*, 9–18.

Wu, Z. (2008). *英语口译笔记法实战指导 [Practical guide of English interpretation note-taking]*. Wuhan: Wuhan University Press.

Xu, H., & Chai, M. (2008). 汉英交替传译活动中译员笔记困难及其原因的实证研究——以国际会议职业受训译员和非职业译员为例 [Difficulties perceived by professional trainee interpreters and non-professional interpreters in note-taking when doing consecutive interpreting: An empirical inquiry through stimulated recall]. *Foreign Language Research*, *1*, 122–127.

Zhong, S. (1999). *A practical handbook of interpretation (enlarged and revised edition)*. Beijing: China Translation and Publishing Corporation.

Zhou, P., Su, X., & Leydesdorff, L. (2010). A comparative study on communication structures of Chinese journals in the social sciences. *Journal of the American society for Information science and Technology*, *61*(7), 1360–1376.

## An introductory note to Chapter 3

As has been made clear in Chapter 2, cognitive load is a core issue in consecutive note-taking. If cognitive load can be measured while interpreters take notes and interpret, the fundamental principles underlying the note-taking choices might be unveiled.

However, the measurement of cognitive load in interpreting is not an easy undertaking. The definition of the construct is somewhat under-specified in interpreting research, accompanied by a lack of systematic discussion on its measurement. Fortunately, the construct of cognitive load has already received ample attention in other fields of research before being introduced to the field of interpreting. Findings in these adjacent fields can inform how cognitive load should be defined and measured in interpreting studies.

Chapter 3 identifies two of the most relevant constructs to the current research: mental workload in human factors research and cognitive load in Cognitive Load Theory research. Building on these two fields, Chapter 3 establishes the theoretical and methodological framework for the PhD study. It provides the validation of an operational definition of the construct of cognitive load in interpreting and discusses its measurement. In particular, this chapter proposes some techniques that are potentially useful for measuring cognitive load in note-taking and CI.

# Chapter 3 The construct of cognitive load in interpreting and its measurement[1]

**Abstract:** Interpreting is a cognitively demanding task, and cognitive load in interpreting is an intriguing topic of research. It is consequently somewhat surprising that relatively little research has been devoted to the topic to date. This article attempts to contribute to that effort by presenting an in-depth discussion on the construct of cognitive load in interpreting and its measurement. Borrowing from mental workload and Cognitive Load Theory research, cognitive load in interpreting is defined as the portion of an interpreter's limited cognitive capacity devoted to performing an interpreting task in a certain environment. The article then presents a methodological discussion on how to measure cognitive load, focusing on the major categories of cognitive load measures and a series of selection criteria. Considering that existing studies only focus on simultaneous interpreting, the article also introduces some techniques that are potentially useful for measuring cognitive load in consecutive interpreting, including the NASA Task Load Index (NASA-TLX), pen recording and eye tracking.

**Keywords:** cognitive load; interpreting; measurement; NASA-TLX; pen recording; eye tracking

## 3.1 Introduction

Interpreting is seen by many scholars as a cognitively demanding task, and many have approached the topic from a cognitive perspective. Such an interdisciplinary effort comes both from outside and within the field. On the one hand, researchers from psychological and psycholinguistic backgrounds (e.g. Barik, 1973; Christoffels, 2004; Christoffels & De Groot, 2004, 2005; De Groot, 1997; Gerver, 1976; Köpke & Signorelli, 2012) are interested in interpreting because it is a special and complex language processing task. They hope that investigating the cognitive processes in interpreting will shed light on how the human mind processes language under severe challenges and in the presence of a high level of multi-tasking. On the other hand, researchers with a background in interpreting (e.g. Moser-Mercer, 1997; Seeber, 2011,

[1] This chapter is a published (on-line first publication) journal article: Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*. doi:10.1080/0907676X.2016.1278026

2013; Seeber & Kerzel, 2012; Shlesinger, 2000) wish to approach the topic from an interdisciplinary perspective. They believe that interpreting research could benefit from the theories and empirical findings in cognitive sciences.

The measurement of cognitive load has received some attention in studies where interpreting is investigated from a cognitive perspective. Most of these studies highlight one particular technique for measuring cognitive load, but methodological discussions on a cohort of measures are scant (an exception to this is Seeber (2013), whose study will be discussed in the following section). However, little effort has been devoted to illustrating the nature of the construct of cognitive load. Moreover, all existing research efforts cater only for simultaneous interpreting, leaving consecutive interpreting largely ignored. It is against this backdrop that the article sets out to initiate an in-depth discussion of the construct of cognitive load and its measurement in interpreting. It is hoped that such a discussion on a theoretical and methodological framework can attract more research interest to the topic and reveal potential directions for future studies.

To that end, the article firstly investigates how cognitive load has been studied in the field of interpreting. Secondly, against the background of the lack of discussion on the nature of the construct of cognitive load, an overview is provided of how similar constructs, i.e. mental workload in human factors research and cognitive load in Cognitive Load Theory, have been defined and investigated, and what can be borrowed from these adjacent fields. The mental workload model by Meshkati (1988) is identified as the most important reference for the current study. Thirdly, the article presents a detailed illustration of cognitive load in interpreting based on the mental workload model. Meshkati’s model is adapted to the specific case of interpreting. Fourthly, a methodological discussion on the main categories of cognitive load measures is presented. The theoretical status of the measures is demonstrated, and a series of selection criteria are discussed. Given the lack of research on cognitive load in consecutive interpreting, an attempt is finally made to introduce a number of techniques that are potentially useful for measuring cognitive load in consecutive interpreting.

The article therefore aims to initiate a discussion on the theoretical and methodological foundation for further empirical research to be carried out on cognitive load in interpreting. The comprehensive illustration of the construct provided in this article will help to standardise the definition and provide an important basis for cognitive load measurement in the field of interpreting.

### 3.2 Investigating cognitive load in interpreting

The interest in cognitive load in interpreting comes both from an attempt to capture and understand the difficult and demanding nature of the task, and from a desire to find out how interpreters deal with the challenges. Few other tasks result in a level of cognitive load similar to that imposed by interpreting, where ‘no physical activity is involved or need be accomplished, no instruments can be of help, everything goes on in the mind’ (Riccardi, Marinuzzi, & Zecchin, 1998, p. 97). Nevertheless, wielding the power of their cognitive systems, interpreters usually succeed in accomplishing this challenging task. Understanding the nature of a construct such as cognitive load in interpreting is important for several reasons. Firstly, it sheds light on the multi-lingual processing that takes place in the human mind under challenging conditions. Secondly, the skills and strategies used by interpreters to cope with the high load are a central component of interpreting competence (Kalina, 2000). Thirdly, cognitive load is indicative of the difficulty of interpreting tasks, which is a key concern in interpreter education and testing, where tasks need to be carefully selected to meet the varied instructional and testing demands (Liu & Chiu, 2009).

Although many researchers have investigated interpreting from a cognitive perspective, cognitive load is studied by only a few, and the main focus is on its measurement. The studies usually apply one particular technique to explicate the cognitive processes in (simultaneous) interpreting or to measure the cognitive load. For example, Petsche, Etlinger, and Filz (1993) found that the electroencephalography (EEG) coherence measure is useful in identifying task-specific cognitive processing. Their experiment with professional interpreters showed that when the verbal task is more difficult, the incidence of coherence increases is higher. In Rinne et al. (2000), brain activation in professional interpreters during simultaneous interpreting was measured by positron emission tomography (PET). It was found that brain activation patterns were modulated by the direction of interpreting. During interpreting into the non-native language, which was deemed the more demanding task, the activation was more extensive. Task-evoked pupillary response (TEPR) is used by Hyönä, Tommola, and Alaja (1995), Seeber and Kerzel (2012), Tommola and Hyönä (1990), and Tommola and Niemi (1986) to measure cognitive load in simultaneous interpreting. In these studies, pupil dilation is found to be indicative of both inter- and intra-task load variations. Furthermore, Gile (2009) and Seeber (2011) use models to assign a priori estimates of cognitive load in interpreting via task analysis.

Despite these pioneering efforts, research conducted on cognitive load in interpreting has been rather limited to date. In particular, the nature of the construct is under-researched. It has been termed variously as mental load, processing load and cognitive load, with these terms used interchangeably in most cases without a formal definition. When it comes to the measurement of cognitive load, the methodological discussions are often insufficient in that they tend to advocate the unique applicability of the one technique being used, but fail to provide a comprehensive picture of the cohort of measures, their respective theoretical statuses, the selection criteria, and how they could be combined.

An exception is Seeber (2013), who has contributed a discussion on all the common measures and methods, their applications in interpreting research, and their respective advantages and disadvantages. It is also the only study in the field to formally define the construct. Seeber (2013) defines cognitive load as 'the amount of capacity the performance of a cognitive task occupies in an inherently capacity-limited system'. This is an exciting step towards a systematic investigation of the topic. However, research on cognitive load in the field is still very limited to date. More efforts are required to get an in-depth understanding of what cognitive load in interpreting is, and to carry out the measurement on a solid foundation.

Fortunately, the construct of cognitive load was not new when it was introduced to interpreting research. Similar constructs have already been investigated, with two of the most relevant ones being mental workload in human factors research, and cognitive load in Cognitive Load Theory. In the following section, we will briefly review how these two constructs are defined and studied in their respective fields, and what can be borrowed to illustrate cognitive load in interpreting.

## 3.3 Mental workload in human factors research and cognitive load in Cognitive Load Theory

Mental workload is a key concern in human factors research, which focuses on how humans accomplish tasks in the context of human-machine system operation, and how different variables affect that accomplishment (Meister, 1989). Investigating the relation between mental workload and human performance is a central research focus, and there has been a long-standing interest in defining and measuring the workload of human operators. Existing definitions of mental workload usually feature slightly different terms, mainly due to the research needs associated with different task types. For example, O'Donnell and Eggemeier (1986) define the construct as 'that portion of the operator's limited capacity actually required to perform a particular task'. Curry, Jex, Levison, and Stassen (1979, p. 236) define the construct as 'the mental effort that the human operator devotes to control and/or supervision relative to his capacity to expend mental effort'.

Cognitive Load Theory, as the name suggests, is a theory that puts cognitive load in the centre. It attempts to explore the effects of instruction on cognitive load, and in turn on learning. Researchers are interested in the cognitive load brought by different instructional methods. The construct has been defined by Paas and van Merriënboer (1994, p. 353) as 'the load that performing a particular task imposes on the cognitive system of a learner'.

What is important to note is that the two constructs both capture the interactions between a particular task and a human with limited cognitive capacity. In that sense, both constructs are relevant to interpreting, where the concept of cognitive load should be able to capture the interactions between an interpreting task (in a certain environment) and an interpreter (with limited cognitive capacity). Therefore, both fields serve as good references for our definition and illustration of cognitive load in interpreting. In this study, we will mainly draw on mental workload studies, while referring to Cognitive Load Theory research where applicable. This is mostly because the latter puts less emphasis on individual characteristics (one admitted limitation of the theory (Bannert, 2002; Moreno & Park, 2010)), whereas interpreter characteristics could potentially affect cognitive load in interpreting in important ways (see Section 4.2).

A representative and influential mental workload model is given by Meshkati (to see a full graphic model please refer to Meshkati (1988, p. 307)). The model consists of a causal section and an effect section, each consisting of two groups of variables. On the causal section are variables that cause mental workload, including 'task and environmental variables' and 'operator's characteristics and moderating variables'; on the effect section are variables affected by mental workload, including 'difficulty, responses and performance' and 'mental workload measures' (Meshkati, 1988, pp. 306-308). Based on the mental workload model, we will now try to define the construct of cognitive load in interpreting, and illustrate the important variables in detail.

## 3.4 Cognitive load in interpreting: an illustration of the construct

Based on the definitions of similar constructs in adjacent fields, we now propose to define cognitive load in interpreting as that portion of an interpreter's limited cognitive capacity devoted to performing an interpreting task in a certain environment. It is a multi-dimensional construct that reflects the interaction of two main groups of variables: task and environmental characteristics on the one hand, and interpreter characteristics on the other hand (Figure 3.1). Task and environmental characteristics determine the amount of mental work to be done in a certain task under certain circumstances. This dimension of cognitive load is sometimes called the "input load" (Johannsen, 1979, p. 4), or "mental load" (Paas & Van Merriënboer, 1994, p. 353). Interpreter characteristics are closely related to the effort that is exerted and experienced by a particular interpreter. This dimension of cognitive load is sometimes called the "operator effort" (Johannsen, 1979, p. 4) or "mental effort" (Paas & Van Merriënboer, 1994, p. 353).

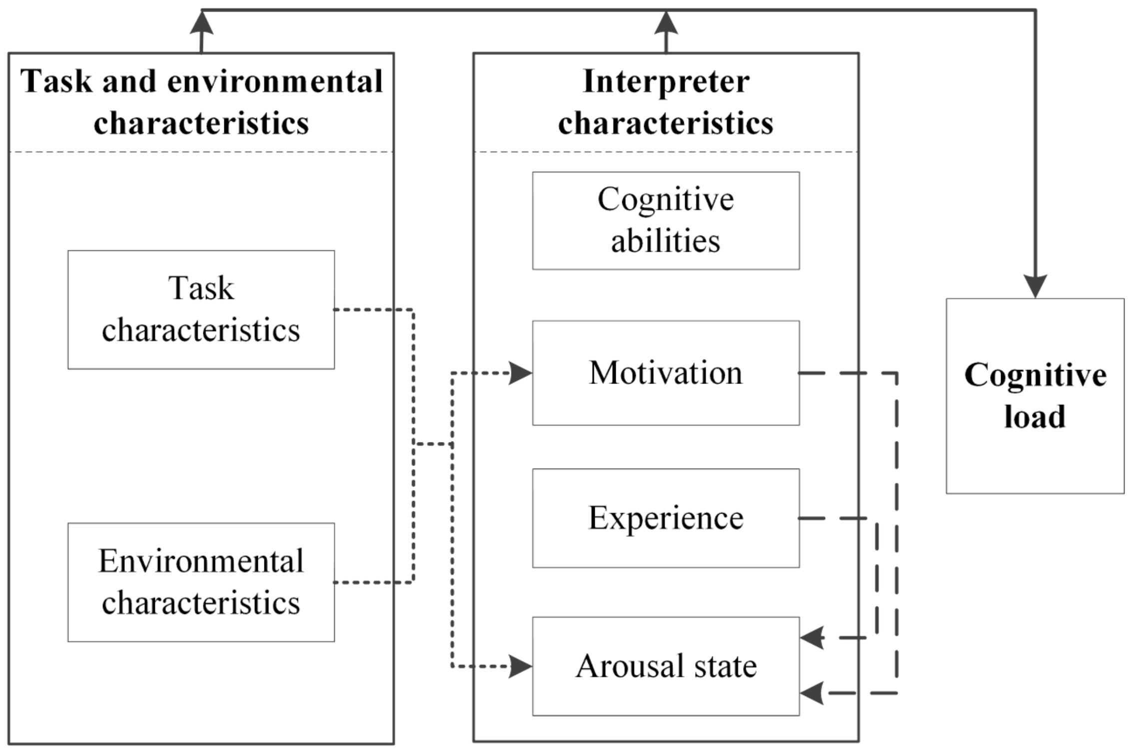


Figure 3.1 A graphical illustration of the construct of cognitive load in interpreting

### 3.4.1 Task and environmental characteristics

#### 3.4.1.1 Task characteristics

Task characteristics is a topic of interest for many interpreting researchers. Some study them from the perspective of testing (e.g. Chao, 2015; Chen, 2009), aiming to build a framework of task characteristics for designing authentic interpreting tests. Others start from the perspective of quality (e.g. Kalina, 2002), looking at factors that could determine the potential quality of a given interpretation. Drawing on those studies, the task characteristics included here are: interpreting mode (e.g. simultaneous or consecutive interpreting), language pair (e.g. some difficulties in interpreting are specific to certain language pairs), interpreting direction (e.g. from or into the native language), features of the speech (formal features such as length, speed, and scripted/spontaneous speech, and content features such as topic, lexical and syntactic complexity), features of the speaker (e.g. native or non-native speaker, accent, speaking competence), expected response (e.g. accuracy, language quality, delivery), time on task (e.g. total hours of working, duration of one turn), preparation (e.g. availability of background material and advance text), task criticality (the level of harm associated with poor task performance), and task novelty (how novel an interpreting task is to an interpreter).

Existing studies on these task characteristics give us a glimpse of the extent to which they could potentially affect cognitive effort in interpreting. For example, Seeber and Kerzel (2012) studied the influence of morphosyntactic asymmetry between source and target language on cognitive load during simultaneous interpreting. They found that when verb-final and verb-initial constructions were interpreted into a verb-initial language like English, the former induced larger pupil dilation, suggesting higher cognitive load. The International Association of Conference Interpreters (AIIC) commissioned a workload study on professional interpreters (AIIC, 2002). According to the interpreters surveyed, many of the factors impacting on the level of burnout ('a combination of physical fatigue, emotional exhaustion and cognitive weariness' (AIIC, 2002, p. 12)) were intrinsic task characteristics, such as long hours in the booth, high-speed speech, and strong accents of the speaker. Moreover, interpreters were asked to rate the frequency of potential difficulties in their work. According to the results, 60% of the interpreters rated the frequency of 'not receiving background material' high, and 44% rated the frequency of 'not having enough time to prepare' high (AIIC, 2002, p. 27). Moser-Mercer, Künzli, and Korac (1998) looked at how increased time on task affected interpreting performance. The study showed that with task speed held steady, stress level increased with time during the first 30 minutes of an interpreting task, and with further time on task (30–60 minutes), performance quality dropped significantly, indicating cognitive overload (see Section 5.2).

Task criticality and task novelty are not often discussed in interpreting research. Task criticality is 'the level of harm associated with performing the task poorly' (Bunch, 2001, p. 1). For example, interpreting for a national leader giving a welcome speech in a diplomatic context is usually considered more critical than interpreting a welcome speech at a business banquet or academic conference. Given the same intrinsic task characteristics, the level of cognitive load associated with the former task is likely to be higher than the latter. Task novelty refers to how novel an interpreting task is to an interpreter. Different tasks vary in novelty, and even the same task could present different levels of novelty to different interpreters, especially considering that many interpreters have their areas of specialisation. For example, interpreters who specialise in business negotiations would find engineering conferences present a higher level of novelty.

#### 3.4.1.2 Environmental characteristics

Environmental characteristics can potentially play important roles in affecting cognitive load in interpreting (a typical example would be the detrimental effect of noise on interpreting performance). A list of these characteristics can also provide a guideline in terms of which variables to control for in order to create better working conditions for interpreters, and to obtain more valid experimental results in research.

Environmental variables that are important to the discussion on cognitive load in interpreting include: physical environment conditions (e.g. location, seating condition, noise, lighting, temperature, air circulation and quality), visibility of the speaker and/or audience, and the equipment used. Poor conditions in the physical environment could lead to an increase in load, as is clearly evident from the results obtained by Parsons (1978). Visibility of the speaker and/or audience is a factor quite unique to the task of interpreting. Its importance in facilitating interpreting has long been asserted by professional interpreters (e.g. Gile, 1990; Rennert, 2008). Although some studies failed to detect a significant difference in performance between interpreting from visual and auditory speeches (Jesse, Vrignaud, Cohen, & Massaro, 2000), it was admitted that the reason could be the presentation of the auditory signal without noise. Moreover, visibility can become especially important 'when the verbal message refers to something visible to the audience or when the nonverbal adds information not present in the verbal message' (Rennert, 2008, p. 204). Equipment used in interpreting is typically not complicated. In simultaneous interpreting, there is usually a control panel, together with earphones and microphones. In consecutive interpreting, the interpreter sometimes relies exclusively on pen and paper. However, there are some situations where slightly more complicated equipment could be used. Typical examples would be various types of remote interpreting, such as telephone and videoconference interpreting. There is also a new form of interpreting called ‘simultaneous consecutive interpreting’ (Hamidi & Pöchhacker, 2007), where the source speech is first recorded and then played back to the interpreter via earphones, and rendered in the simultaneous mode. Under such circumstances, the influence of equipment on cognitive load could be more significant.

### 3.4.2 Interpreter characteristics

Interpreter characteristics that affect cognitive load most significantly include the cognitive abilities, motivation, experience, and state of arousal or activation level of the interpreter.

Cognitive abilities are at the heart of discussions on such topics as interpreter competence, expertise and aptitude (e.g. Hoffman, 1997; Macnamara, 2012). Cognitive abilities that have a strong influence on cognitive load in interpreting include not only general abilities such as intellect, knowledge (both general knowledge and topical knowledge), language proficiency, cultural competence, and memory (especially working memory), but also skills that are specific to certain types of interpreting, such as note-taking in consecutive interpreting.

Motivation is closely related to an interpreter’s goals and attitude towards a task, and can affect the focus and level of effort expended on the task. Although there is a lack of research on its impact on cognitive load, the importance of motivation as a determinant of interpreting performance has been discussed in multiple studies, especially from the perspectives of skill acquisition and training (e.g. Moser-Mercer, 2008; Timarová & Salaets, 2011).

The experience of an interpreter comes from both training and working. Experience is a known variable to influence mental workload (O'Donnell & Eggemeier, 1986; Young & Stanton, 2001). However, the mechanism of how experience and cognitive load interact in interpreting remains unclear due to a lack of research. To investigate the issue, a good starting point might be a comparison between the performance of interpreters with varied levels of experience (e.g. Köpke & Nespoulous, 2006; Liu, Schallert, & Carroll, 2004).

State of arousal refers to the physical activation level of an interpreter. To achieve an optimum level of task performance, it is necessary to have a certain level of stimulation or arousal. This level of stimulation or arousal varies from interpreter to interpreter. One simple example would be: some interpreters might find themselves functioning better in the morning (attentional mechanisms more active), while others might be more efficient in the afternoon.

### 3.4.3 Interactions

Interactions (marked by dotted lines in Figure 3.1) could happen both between and within the two groups of variables in the causal dimension. Between the groups, firstly, the state of arousal is affected by task criticality, task novelty, and environmental factors. Higher levels of task criticality and novelty could lead to a higher level of arousal. Environmental hazards could also lead to high arousal levels which are harmful to task performance, increasing the risk of cognitive overload. Secondly, task criticality affects motivation. Increased task criticality could motivate an interpreter, putting the interpreter in a better state to marshal cognitive resources.

Interactions within the group of interpreter characteristics are shown in that both motivation and experience affect an interpreter's arousal state. Motivation usually leads to higher arousal levels. A lack of experience could also lead to higher arousal, especially compared to situations when an interpreter is too familiar with a task and even feels bored.

So far, cognitive load in interpreting has been conceptualised as a multi-dimensional construct reflecting the interactions between task and environmental characteristics and interpreter characteristics. On this basis, the following section presents a methodological discussion on how cognitive load in interpreting could be measured by discussing the major categories of measures and introducing a series of selection criteria.

## 3.5 Measuring cognitive load in interpreting: a methodological discussion

### 3.5.1 Cognitive load measures

Since cognitive load is essentially a theoretical construct, it cannot be observed and measured directly. What we can do is to rely on observable and measurable surrogates that are indicative of cognitive load. A seemingly obvious indicator is the interpreter's subjective feeling of effort. The assumption is that, with increased capacity expenditure, the interpreter would feel effort or exertion, which could be self-evaluated with a rating scale. A second indicator is the interpreting performance. The rationale is that a decrease in the quality of performance (evident from an increase in elements such as errors, omissions, and pauses) is likely to be associated with an increase in cognitive load. A potential third indicator is the physiological arousal of the interpreter. The assumption is that effort, a major determinant of cognitive load (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), is quantifiable through measuring the activation level of the human body. A fourth possible indicator is the interpreting task characteristics. This is an a priori estimate of cognitive load by analysing task complexity.

The four types of indicators are associated with four categories of cognitive load measures: subjective measures, performance measures, physiological measures, and analytical measures (Figure 3.2). Subjective measures are usually produced using psychometric rating scales. The scales can be either unidimensional or multidimensional. Unidimensional scales treat cognitive load as a unitary construct and the subject must assign a single rating to characterise the exerted effort. Multidimensional scales reflect several factors that contribute to the subjective feelings of effort expenditure and allow separate ratings on each factor. Performance measures include two types: primary task measures and secondary task measures. Primary task measures use the interpreting performance to indicate cognitive load changes. Secondary task measures are produced through the concurrent performance of an interpreting task and an additional task (the secondary task). Changes in the performance of the secondary task are evaluated as evidence of the available spare capacity. Physiological measures approach cognitive load by observing functions of different body parts such as brain, eye, cardiac system and muscle. Analytical measures are usually provided by experts or derived from models or task analysis, based on current knowledge about the task (Paas et al., 2003).

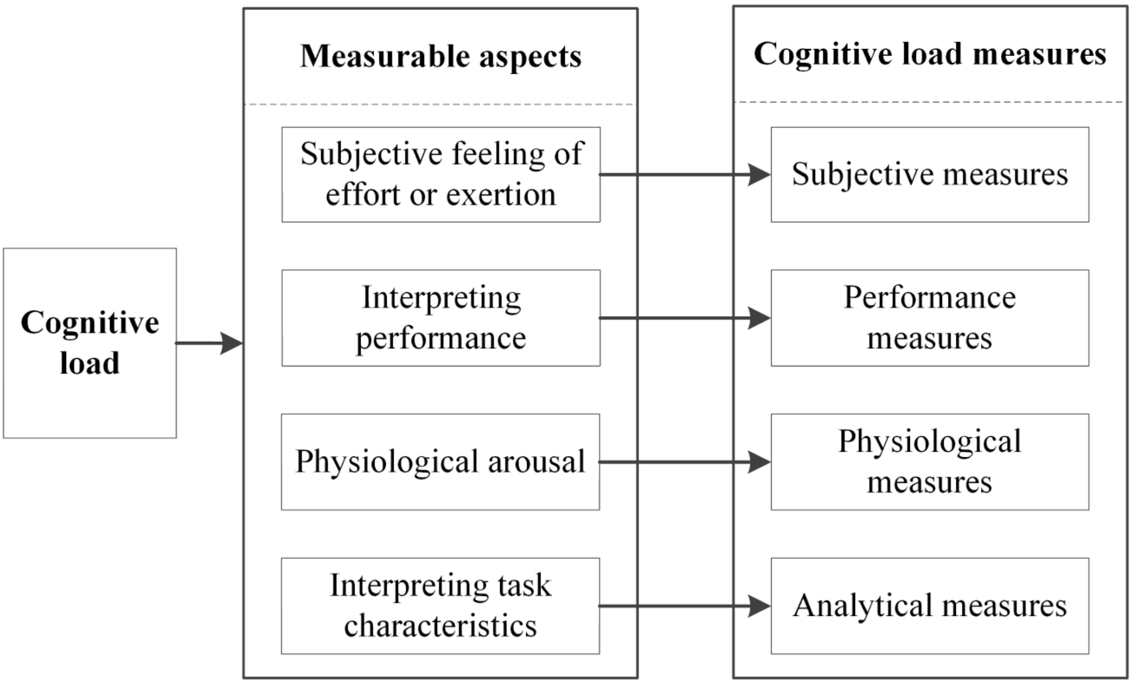


Figure 3.2 A graphical illustration of the measurement of cognitive load in interpreting

The cognitive load measures and their application examples both outside and within the field of interpreting are summarised in Table 3.1.

Table 3.1 Major categories of cognitive load measures and their applications

| **Measures** | | **Representative examples in adjacent fields** | **Pioneering studies in the field of interpreting** |
|---|---|---|---|
| Subjective measures | Subjective rating scales | Hart and Staveland (1988) | |
| Performance measures | Primary task measures | Paas et al. (2003) | Moser-Mercer et al. (1998) |
| | Secondary task measures | Brünken, Steinbacher, Plass, and Leutner (2002) | Hu (2008) |
| Physiological measures | Brain measures | Anderson et al. (2011) | Petsche et al. (1993); Rinne et al. (2000) |
| | Eye measures | Beatty (1982) | Tommola and Niemi (1986); Tommola and Hyönä (1990); Hyönä et al. (1995); Seeber and Kerzel (2012) |
| | Cardiac system measures | Gunn, Wolf, Block, and Person (1972) | Klonowicz (1994) |
| | Muscle measures | Leyman, Mirka, Kaber, and Sommerich (2004) | |
| Analytical measures | Expert opinion | Kuperman (1985) | |
| | Models | Wickens (2002) | Gile (1995/2009); Seeber (2011) |
| | Task analysis | Sweller (1988) | |

Because cognitive load is a multi-dimensional construct, a single measure cannot provide a comprehensive picture. Moreover, the measures differ in their granularity and their relation to the interpreting event in time (i.e. real-time vs. post hoc). Subjective measures can only provide a post hoc and overall indication of cognitive load. Performance measures can offer real-time indicators of cognitive load, but they are only sensitive when the level of load begins to exceed the capacity of the interpreter to compensate. Physiological measures are both real-time and objective, but their accurate interpretation is usually reliant on additional subjective measures. Analytical measures can only be used to estimate the input dimension of cognitive load. Given that the measures have different strengths and weaknesses, researchers need to consider the criteria for selecting the appropriate measures for different research situations and purposes.

### 3.5.2 Selection criteria

A number of studies have discussed the criteria for selecting mental workload measures. Two representative examples are O'Donnell and Eggemeier (1986) and Wickens and Hollands (1999). Among the criteria that have been proposed, the following are especially relevant to interpreting research: sensitivity, diagnosticity, and intrusiveness.

*Sensitivity* refers to the potential of a measure to discriminate between changes in cognitive load. To determine the sensitivity of a measure, we will need to refer to the theoretical relationship between performance and cognitive load (Figure 3.3, adapted from O'Donnell and Eggemeier (1986)). The curve specifies three regions according to the level of cognitive load. In region A, the interpreter has sufficient capacity to cope with the increasing cognitive load without sacrificing the quality of interpreting performance. Since no changes could be observed in performance in this region, primary task measures would be insensitive, while subjective and physiological measures are more suitable. In region B, cognitive load begins to exceed the capacity of the interpreter to compensate, and performance decreases and becomes sensitive to load changes. This is where cognitive load can be, and usually is, measured by primary task performance. In region C, the level of load is too high, and performance drops to a catastrophic level. All measures would indicate high cognitive load, but it would be difficult to differentiate the levels of load. In practice, research is usually centred on measuring cognitive load in regions A and B.

*Diagnosticity* refers to the potential of a measure to identify the specific cause of cognitive load from a variety of sources in the cognitive system. It is based on Wickens's (2002) Multiple Resources Theory, which assumes that there are more than one reservoirs of resources within the human processing system. Sometimes it is necessary to identify which resources are utilised or to differentiate the level of demands on certain resources. Some measures (e.g. pupillometry and subjective measures) only indicate the overall load level and are therefore not considered diagnostic. Other more diagnostic measures (e.g. secondary task measures and event-related brain measures) could be used to measure demands on certain resources, but they are insensitive to other resource types.

*Intrusiveness* is determined by the extent to which a measure interrupts the performance of the interpreting task. This is an important concern for interpreting research because the cognitive system of interpreters is subject to heavy load during the task. Measures that pose minimum intrusion on the interpreting performance, such as subjective rating scales, primary task measures, some low-invasive physiological measures, and analytical measures, are more suitable for interpreting research. Secondary task measures, however, should be treated with extra caution because of their intrusiveness.

These criteria should be carefully considered when choosing techniques to measure cognitive load in interpreting. They are related not only to specific research purposes but also to the type of interpreting task being investigated. For simultaneous interpreting, the usefulness of certain techniques has been demonstrated (see Section 2). For consecutive interpreting, however, discussions on measuring cognitive load are very rare. Although some techniques could be used for both modes of interpreting, the unique features of consecutive interpreting present new challenges as well as potentials. Based on our methodological discussion, the remaining part of this article will be dedicated to proposing some techniques that are potentially useful for measuring cognitive load in consecutive interpreting.

## 3.6 Techniques for cognitive load measurement in consecutive interpreting

### 3.6.1 Subjective rating scales

There is little evidence that subjective rating scales have been employed to measure cognitive load in interpreting, but they have been used widely in mental workload and

Cognitive Load Theory research, and have been proven valid, non-intrusive, and easy to implement. It would, therefore, be meaningful to determine the usefulness of some established rating scales from these areas to interpreting research.

The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is one of the most widely used scales for measuring mental load. It is a multi-dimensional rating procedure that provides an overall load score based on a weighted average of ratings on six subscales: mental demands, physical demands, temporal demands, performance, effort, and frustration. Participants need to complete pair-wise comparisons on all subscales and indicate which is more relevant to their personal load definition. The number of times a subscale is chosen is the weight. This weighting scheme is used to take individual differences into account. Participants also need to rate on each subscale by giving a score that best represents the load experienced during task performance. The NASA-TLX could be used for both simultaneous and consecutive interpreting. The rating should be done immediately after an interpreting task.

The NASA-TLX is traditionally done with pen and paper, but now there are also a few computerised versions available (e.g. Cao, Chintamani, Pandya, & Ellis, 2009). The digital versions are much faster and more convenient to use. In an exploratory study run by the author with five professional interpreters, the average time needed to complete the computerised NASA-TLX[2] is under ten minutes. Some studies propose and apply modified versions of the original scale by adding, deleting or redefining the existing subscales to improve their relevance to specific tasks or experimental questions (Hart, 2006). This strategy could also be used to better fit the NASA-TLX to interpreting research, but the validity, sensitivity, and reliability of the modified instrument need to be tested with empirical data.

### 3.6.2 Primary task performance, pen recording, and eye tracking

Performance has been a prime concern of interpreting research over the years. According to the selection criterion of sensitivity (see Section 5.2), primary task performance can serve as an objective indicator of cognitive load when load begins to exceed the capacity of the interpreter to compensate, and performance decreases and becomes sensitive to load changes (region B in Figure 3.3). Under such circumstances, a performance score could indicate the overall cognitive load induced by a certain task. Performance scores are therefore sometimes used to determine the difficulty of

[2] A link to this computerised version of the NASA-TLX can be found at http://www.nasatlx.com/.

interpreting tasks (usually in combination with subjective reports and/or expert opinion) in interpreter education and testing (e.g. Liu & Chiu, 2009).

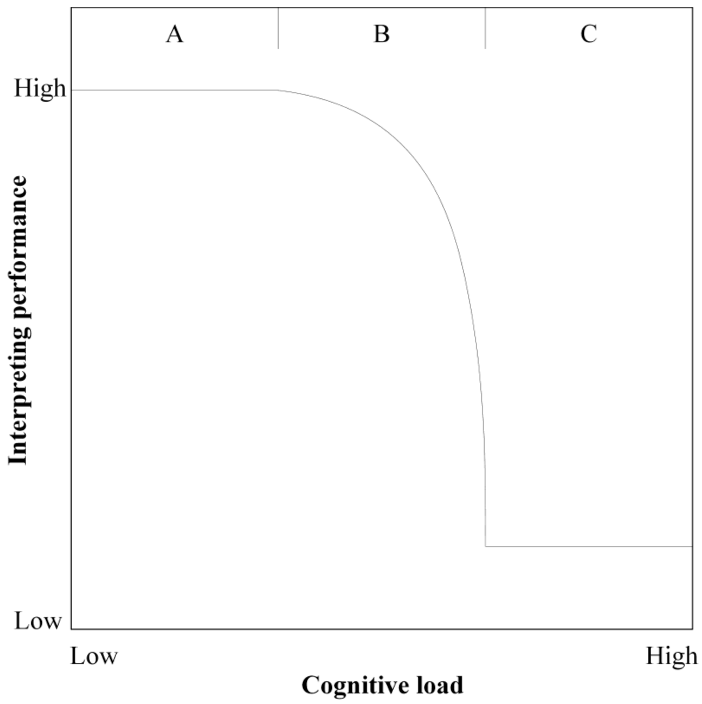


Figure 3.3 A hypothetical relationship between cognitive load and interpreting performance

Other performance measures such as the target speech features of span and errors have also been studied by many. For example, pauses in consecutive interpreting have been studied by Mead (2000, 2002, 2005), while ear-voice-span and errors in simultaneous interpreting have been studied by Altman (1994), Barik (1973), and Bendazzoli, Sandrelli, and Russo (2011). Although the target speech features are not usually viewed from a cognitive load perspective, those initial investigations have laid a useful foundation by presenting methods to analyse and quantify interpreting performance. Now with physiological measures (such as eye tracking) gaining popularity, performance indicators could be collected in ways that have not been possible before, providing more powerful tools to reveal information on cognitive load. Two potential indicators that can be investigated by combining performance and physiological techniques, one for each of the two phases of consecutive interpreting, deserve particular attention.

Phase I of consecutive interpreting is the ‘listening and note-taking phase’ (Gile, 2009, p. 175). The indicator that can be investigated in this phase is the ear-pen span, which is the interval between the moment a speech segment is heard and the moment it is noted down. More specifically, it is calculated from the offset of voice to the onset of pen stroke. The hypothesis is that longer ear-pen span indicates higher cognitive load. When a source speech unit is difficult and the cognitive load is high, it takes longer to process that unit, to make the decision, and to put that unit into written notes. But a long span could also indicate high cognitive load in processing units other than the one that is noted down. A more fine-grained analysis could be reached by combining the span data with retrospection. The recorded phase I can be played back to the interpreter, providing retrieval cues to ensure accurate and comprehensive retrospection.

Using a piece of software called the Eye and Pen[3], the note-taking process taking place on a digital tablet can be recorded. The tablet works together with a digital pen, and transmits the spatial, temporal and pressure data to the software as the pen moves across its surface. The software can then analyse and reconstruct the writing process, giving real-time data such as the distance, duration, speed, and pressure of the pen, as well as the pauses between pen strokes. The source speech and the pen data share the same timeline, so it is possible to accurately pinpoint the offset of any source speech unit and the onset of any pen stroke, allowing the calculation of the ear-pen span.

Phase II of consecutive interpreting is the ‘speech production phase’ (Gile, 2009, p. 176). The indicator that can be investigated in this phase is the eye-voice span[4], which is the interval between the moment a note unit is read and the moment it is produced in the target speech. More specifically, it is the interval between the onset of a note’s fixation and the onset of its articulation. The hypothesis is that longer eye-voice span indicates higher cognitive load. If a note unit is difficult to process (for example when a highly abstract symbol is used or the handwriting is illegible) and the associated cognitive load is high, it takes longer before the interpreter can produce an equivalent in the target speech. Again, a retrospection of the production phase (possibly stimulated by the scan path video of eye movements) can help with a more fine-grained analysis of the span data. For example, it can help distinguish whether a long eye-voice span is caused by the difficulty in processing a certain note unit, or by recalling from memory information that’s not relevant to the note unit.

---

[3] The website of the software is http://www.eyeandpen.net/?lng=en.

[4] This is different from the eye-voice span widely discussed in reading research. The eye-voice span in reading research is measured as the distance between the fixated item and the pronounced item during reading aloud.

Using eye tracking, the production phase can be recorded in detail. Eye tracking data reveals information on when and for how long a note unit is being fixated. Some eye trackers (such as the SMI Eye Tracking Glasses and the Tobii Pro Glasses) could synchronously collect eye data while recording a scene video with sounds. The eye data and the voice data share the same timeline, making it possible to calculate the eye-voice span.

Of course both pen recording and eye tracking yield other potential indicators of cognitive load, such as pen orientation and pressure (e.g. Yu, Epps, & Chen, 2011a; Yu, Epps, & Chen, 2011b) and number and duration of eye fixations (e.g. Buettner, 2013; Just & Carpenter, 1976, 1980; Van Orden, Limbert, & Makeig, 2001). Different sources of data could be triangulated to present a comprehensive picture of cognitive load in consecutive interpreting.

### 3.6.3 Models

The analytical measures of expert opinion and task analysis are used mostly in interpreting education and testing for the purpose of gauging task difficulty. Teachers and test developers often have to deal with tasks that vary in different aspects. They make judgements about the complexity, or difficulty, of the tasks based on those different aspects in order to choose appropriate practising materials and test tasks to match different training and testing objectives (Liu & Chiu, 2009). Models, compared with the previous two types of analytical measures, are used more for research purposes. Two outstanding examples are Gile's (2009) Effort Models, and Seeber's (2011) Cognitive Load Models. The latter is especially suitable for making predictions on local cognitive load.

The Cognitive Load Models developed by Seeber are based on Wickens's (2002) Multiple Resources Theory (see Section 5.2). They are able to quantify cognitive load relying principally on Wickens's demand vectors and conflict coefficients. Local cognitive load changes are reflected by analysing the trade-offs between time-sharing activities in the interpreting process. Although the models are designed for simultaneous interpreting, they can be adapted to cater to the situations in consecutive interpreting. Using the same principles, the first phase of consecutive interpreting can be considered as a real-time combination of a listening and a note-taking task, while the second phase consists of a note-reading and a speech production task. Both tasks can be broken down into their resource demand vectors and analysed using the Multiple Resources Theory.

The adapted models will be able to make specific predictions on the changes in cognitive load brought by different strategies and serve as a useful analytical technique for estimating cognitive load in consecutive interpreting.

It exceeds the scope of this article to expand the introduced techniques in full or to present detailed application procedures. This article is only an initial step in a more comprehensive research project on cognitive processing and cognitive load in consecutive interpreting. However, it is hoped that the brief introductions made here of the potential techniques set the agenda for more comprehensive and meticulous experimental research on cognitive load during consecutive interpreting.

### 3.7 Conclusions

Building on previous studies in the field of interpreting, and borrowing from adjacent fields such as mental workload and Cognitive Load Theory research, this article attempts to formally define and illustrate the construct of cognitive load in interpreting and to discuss its measurement. Cognitive load in interpreting is conceptualised as a multi-dimensional construct that reflects the interaction between an interpreting task (in a certain environment) and an interpreter (with certain characteristics). It is a theoretical construct that cannot be observed directly, and its measurement relies on observable surrogates that are indicative of cognitive load. There are four major categories of cognitive load measures: subjective, performance, physiological and analytical measures. They cater to different research purposes and are applicable in different circumstances. A series of criteria that could help with the selection and combination of the measures is also discussed. Finally, considering the lack of research on cognitive load in consecutive interpreting, the article introduces some techniques that are potentially useful for measuring cognitive load in consecutive interpreting. Our next step is to validate the measurement techniques introduced in this article. Although they have rarely been applied in empirical research so far, the techniques have provided some exciting possibilities.

The significance of defining and operationalising cognitive load in interpreting lies in the foundation it provides for investigating the cognitively challenging task of interpreting experimentally. Such a theoretical framework can be helpful in identifying triggers of cognitive overload, facilitating research on the interpreting process. Research on cognitive load in interpreting is only starting, and hopefully more interested researchers will be joining the effort.

## 3.8 References


AIIC. (2002). Interpreter workload study - Full report. Retrieved 7 February 2002 http://aiic.net/page/657

Altman, J. (1994). Error analysis in the teaching of simultaneous interpretation: A pilot study. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 25–38). Amsterdam: John Benjamins.

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum, 30*(3), 791–800. doi:10.1111/j.1467-8659.2011.01928.x

Bannert, M. (2002). Managing cognitive load: Recent trends in cognitive load theory. *Learning and Instruction, 12*, 139–146.

Barik, H. C. (1973). Simultaneous interpretation: Temporal and quantitative data. *Language and Speech, 16*(3), 237–270. doi:10.1177/002383097301600307

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276–292.

Bendazzoli, C., Sandrelli, A., & Russo, M. (2011). Disfluencies in simultaneous interpreting: A corpus-based analysis. In A. Kruger, K. Wallmach, & J. Munday (Eds.), *Corpus-based translation studies: Research and applications* (pp. 282–306). London and New York: Continuum.

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*(2), 109–119. doi:10.1027//1618-3169.49.2.109

Buettner, R. (2013). Cognitive workload of humans using artificial intelligence systems: Towards objective measurement applying eye-tracking technology. In I. J. Timm & M. Thimm (Eds.), KI 2013: Advances in artificial intelligence (pp. 37–48). Berlin Heidelberg: Springer. doi:10.1007/978-3-642-40942-4

Bunch, M. B. (2001). *The SSTI/NAJIT translation and interpretation national certification examination*. Paper presented at the Annual Meeting of the American Translators Association, Los Angeles. http://www.najit.org/certification/ATAPRESR%20on%20Exam%20History%20and%20development%20dlm.PDF

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*(1), 113–117. doi:10.3758/BRM.41.1.113

Chao, H. (2015). *Building the validity foundation for interpreter certification performance testing* (Unpublished doctoral thesis), Macquarie University, Sydney.

Chen, J. (2009). Authenticity in accreditation tests for interpreters in China. *The Interpreter and Translator Trainer, 3*(2), 257–273. doi:10.1080/1750399X.2009.10798791

Christoffels, I. K. (2004). *Cognitive studies in simultaneous interpreting* (Unpublished doctoral thesis), University of Amsterdam, Amsterdam.

Christoffels, I. K., & De Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition, 7*(3), 227–240. doi:10.1017/S1366728904001609

Christoffels, I. K., & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). New York: Oxford University Press.

Curry, R., Jex, H., Levison, W., & Stassen, H. (1979). Final report of control engineering group. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 235–253). New York: Plenum Press.

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation: Three approaches. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 25–56). London and New Delhi: Thousand Oaks.

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. B. W. Anderson & R. W. Brislin (Eds.), *Translation: Applications and research* (pp. 165–207). New York: Gardner Press.

Gile, D. (1990). Scientific research vs. personal theories in the investigation of interpretation. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research on conference interpretation* (pp. 28–41). Udine: Campanotto Editore.

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

Gunn, C. G., Wolf, S., Block, R. T., & Person, R. J. (1972). Psychophysiology of the cardiovascular system. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology*. New York: Holt, Rinehart and Winston.

Hamidi, M., & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta, 52*(2), 276-289. doi:10.7202/016070ar

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B.V.

Hoffman, R. R. (1997). The cognitive psychology of expertise and the domain of interpreting. *Interpreting, 2*(1), 189-230.

Hu, L. (2008). *Note-taking in consecutive interpreting: A psycholinguistic perspective* (Unpublished doctoral thesis), Shanghai International Studies University, Shanghai.

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

Jesse, A., Vrignaud, N., Cohen, M. M., & Massaro, D. W. (2000). The processing of information from multiple sources in simultaneous interpreting. *Interpreting, 5*(2), 95-115. doi:10.1075/intp.5.2.04jes

Johannsen, G. (1979). Workload and workload measurement. In N. Moray (Ed.), *Mental workload: Its theory and measurement* (pp. 3-11). New York: Plenum Press.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441-480. doi:10.1016/0010-0285(76)90015-3

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

Kalina, S. (2000). Interpreting competences as a basis and a goal for teaching. *The Interpreters' Newsletter, 10*, 3-32. Retrieved from http://hdl.handle.net/10077/2440

Kalina, S. (2002). Quality in interpreting and its prerequisites: A framework for a comprehensive view. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 121-130). Amsterdam: John Benjamins.

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

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting, 8*(1), 1-23. doi:10.1075/intp.8.1.02kop

Köpke, B., & Signorelli, T. M. (2012). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism, 16*(2), 183-197. doi:10.1177/1367006911402981

Kuperman, G. G. (1985). Pro-SWAT applied to advanced helicopter crewstation concepts. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 29*(4), 398-402. doi:10.1177/154193128502900420

Leyman, E. L. C., Mirka, G. A., Kaber, D. B., & Sommerich, C. M. (2004). Cervicobrachial muscle response to cognitive load in a dual-task scenario. *Ergonomics, 47*(6), 625-645. doi:10.1080/00140130310001629766

Liu, M., & Chiu, Y.-H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting, 11*(2), 244–266. doi:10.1075/intp.11.2.07liu

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

Macnamara, B. (2012). Interpreter cognitive aptitudes. *Journal of Interpretation, 19*(1), 9–32. Retrieved from http://digitalcommons.unf.edu/joi/vol19/iss1/1

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter, 10*, 89–102. Retrieved from http://hdl.handle.net/10077/2451

Mead, P. (2002). Exploring hesitation in consecutive interpreting: An empirical study. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 73–82). Amsterdam: John Benjamins.

Mead, P. (2005). Methodological issues in the study of interpreters' fluency. *The Interpreters' Newsletter, 13*, 39–63. Retrieved from http://hdl.handle.net/10077/2469

Meister, D. (1989). *Conceptual aspects of human factors*. Baltimore: Johns Hopkins University Press.

Meshkati, N. (1988). Toward development of a cohesive model of workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 305–314).

Moreno, R., & Park, B. (2010). Cognitive Load Theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 9–28). New York: Cambridge University Press.

Moser-Mercer, B. (1997). Beyond curiosity: Can interpreting research meet the challenge? In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 176–195). London and New Delhi: Thousand Oaks.

Moser-Mercer, B. (2008). Skill acquisition in interpreting. *The Interpreter and Translator Trainer*, *2*(1), 1–28. doi:10.1080/1750399X.2008.10798764

Moser-Mercer, B., Künzli, A., & Korac, M. (1998). Prolonged turns in interpreting: Effects on quality, physiological and psychological stress (Pilot study). *Interpreting*, *3*(1), 47–64.

O'Donnell, C. R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, vol. II: Cognitive processes and performance* (pp. 42-41–42-49). New York: Wiley-Interscience.

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. doi:10.1207/S15326985EP3801_8

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371. doi:10.1007/BF02213420

Parsons, H. M. (1978). Human factors approach to simultaneous interpretation. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 315–321). New York: Springer US.

Petsche, H., Etlinger, S. C., & Filz, O. (1993). Brain electrical mechanisms of bilingual speech management: An initial investigation. *Electroencephalography and Clinical Neurophysiology, 86*(6), 385–394. doi:10.1016/0013-4694(93)90134-H

Rennert, S. (2008). Visual input in simultaneous interpreting. *Meta, 53*(1), 204–217. doi:10.7202/017983ar

Riccardi, A., Marinuzzi, G., & Zecchin, S. (1998). Interpretation and stress. *The Interpreters' Newsletter, 8*, 93–106. Retrieved from http://hdl.handle.net/10077/2493

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., . . . Sunnari, M. (2000). The translating brain: Cerebral activation patterns during simultaneous interpreting. *Neuroscience Letters, 294*(2), 85–88. doi:10.1016/S0304-3940(00)01540-8

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting, 13*(2), 176–204. doi:10.1075/intp.13.2.02see

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target, 25*(1), 18–32. doi:10.1075/target.25.1.03see

Seeber, K. G., & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism, 16*(2), 228–242. doi:10.1177/1367006911402982

Shlesinger, M. (2000). Interpreting as a cognitive process: How can we know what really happens? In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research* (pp. 3–16). Amsterdam and Philadelphia: John Benjamins.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285. doi:10.1207/s15516709cog1202_4

Timarová, Š., & Salaets, H. (2011). Learning styles, motivation and cognitive flexibility in interpreter training: Self-selection and aptitude. *Interpreting, 13*(1), 31–52. doi:10.1075/intp.13.1.03tim

Tommola, J., & Hyönä, Jukka. (1990). Mental load in listening, speech shadowing and simultaneous interpreting: A pupillometric study. In J. Tommola (Ed.), *Foreign language comprehension and production* (pp. 179–188). Turku: AfinLA.

Tommola, J., & Niemi, P. (1986). Mental load in simultaneous interpreting: An on-line pilot study. In L. S. Evensen (Ed.), *Nordic research in text linguistics and discourse analysis* (pp. 171–184). Trondheim: Tapir.

Van Orden, K. F., Limbert, W., & Makeig, S. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*(1), 111–121.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. doi:10.1080/14639220210123806

Wickens, C. D., & Hollands, J. G. (1999). *Engineering psychology and human performance (3rd edition)*. Upper Saddle River, New Jersey: Prentice Hall.

Young, M., & Stanton, N. (2001). Mental workload: theory, measurement, and application. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (Vol. 1, pp. 507–509). London: Taylor and Francis.

Yu, K., Epps, J., & Chen, F. (2011a). *Cognitive load evaluation of handwriting using stroke-level features*. Paper presented at the 16th International Conference on Intelligent User Interfaces, Palo Alto, CA, USA. http://www.nicta.com.au/pub-download/full/5228/

Yu, K., Epps, J., & Chen, F. (2011b). *Cognitive load measurement with pen orientation and pressure*. Paper presented at the ICMI 2011 Workshop on Inferring Cognitive and Emotional States from Multimodal Measures, Alicante, Spain. http://www.nicta.com.au/pub-download/full/5228/

## An introductory note to Chapter 4

The theoretical and methodological discussions in Chapter 3 have identified a series of techniques that are potentially useful for measuring cognitive load in note-taking and CI. But before these techniques can be applied to the main study of this PhD project, their applicability needs to be tested, especially the ones that are seldom used in interpreting research.

Chapter 4 reports on a pilot study which focuses on the synchronised recording and analysis of pen and voice data. It documents how the tasks are carefully created and controlled for variance, how the sample participants are selected through stringent criteria, how the apparatus are set up so that abundant and varied sources empirical data can be recorded in parallel, and how data processing and analysis are carried out. The purpose of the pilot study is three-fold: (1) to prove that pen recording is a useful method to tap into the process of note-taking and CI; (2) to devise some useful indicators of cognitive load with the combined recording of pen and voice data; and (3) to provide instrumental findings for the hypotheses to be made in the main study.

# Chapter 4 Note-taking in consecutive interpreting: New data from pen recording[1]

**Abstract:** Note-taking provides a unique opportunity to investigate consecutive interpreting (CI). This study approaches note-taking from a cognitive perspective, combining product analysis with the process research method of pen recording. It investigates such variables as the choice of form, the choice of language, the relationship between note-taking and interpreting performance, and the relationship between note-taking and cognitive load in CI. In the context of CI between Chinese and English, the study finds that interpreters prefer language to symbol, abbreviation to full word, and English to Chinese regardless of the direction of interpreting. Interpreting performance is not directly related to either the quantity or the quality of notes; it is a function of both. Pen recording appears to be a powerful method to tap into the process of note-taking and CI, and the collected data could potentially serve as useful indicators of cognitive load.
**Keywords:** consecutive interpreting; note-taking; pen recording; cognitive load

## 4.1 Introduction

The research interest in cognitive processing in translation and interpreting is increasing, but the focus on consecutive interpreting (CI) is very limited to date. Note-taking is a distinctive feature of CI[2], and provides a unique opportunity to investigate the interpreting process. For over half a century, research on note-taking in CI has yielded fruitful results. A series of variables have been investigated, including the choice of form, the choice of language, and the relationship between note-taking and interpreting performance. However, existing studies on note-taking and CI are mostly product-oriented, revealing little information about the process.

This study attempts to address that limitation by combining product analysis with an investigation into the interpreting process. Using pen recording and a software called the *Eye and Pen*[3], pen data during the note-taking process are recorded in great details.

[1] This chapter is a published journal article: Chen, S. (2017). Note-taking in consecutive interpreting: New data from pen recording. *Translation and Interpreting, 9*(1), 4–23. doi:10.12807/ti.109201.2017.a02

[2] In this article, CI refers to long consecutive where systematic note-taking is used.

[3] The website of the software is http://www.eyeandpen.net/?lng=en.

Pen strokes are measured in terms of distance, duration, and speed. Such a recording not only tells us what interpreters' note-taking choices are, but also shows us how interpreters carry out those choices. The pen data are further investigated from a cognitive perspective, with an aim to see if they can be used as indicators of cognitive load in note-taking and CI.

### 4.2 Note-taking in CI: a brief review

The large volume of literature generated by scholars' sustained interest in note-taking can be roughly divided into two streams: a prescriptive stream and a descriptive stream (see Chen (2016) for a more comprehensive review). At the earliest stage, a number of prescriptive works have introduced some well-known note-taking systems and principles (e.g., Kirchhoff, 1979; Matyssek, 1989; Rozan, 1956/2002). Later on, noticing the challenges brought by the teaching and learning of note-taking in classrooms, some scholars begin to observe how notes are actually taken by student interpreters (e.g., Alexieva, 1994; Gile, 1991). These studies represent the beginning of a shift in note-taking literature from being prescriptive to becoming descriptive. Some researchers have also investigated the cognitive and linguistic aspects of note-taking, pointing out a concurrent storage of information in memory and in notes (e.g., Seleskovitch, 1975) and that note-taking operates on a micro-level that stays close to the source text (e.g., Albl-Mikasa, 2008; Kohn & Albl-Mikasa, 2002). The more recent studies in the descriptive stream usually target specific note-taking choices, collecting data in simulated interpreting tasks and contributing valuable empirical evidence (e.g., Abuín González, 2012; Andres, 2002; Dam, 2004b; Szabó, 2006). In all these studies, three variables have received the majority of the attention: the choice of form, the choice of language, and the relationship between note-taking and interpreting performance.

Interpreters make choices (although not always consciously) on the form of notes: whether to take notes in symbol or language, and if in language, whether to write the word in full or to abbreviate it. Many prescriptive publications introducing note-taking systems put the use of symbols and abbreviations at a prominent position. Compared to language, symbols are easy to write and read, and can help avoid source language influence because they represent concepts rather than specific words (Gillies, 2005, p. 99). But the prescriptive suggestion on how many symbols should be used varies from system to system. At the minimalist end was Rozan, who recommended a total of 20 symbols, of which "only 10 were indispensable" (1956/2002, p. 25). At the maximalist end was Matyssek (1989), who used a whole book volume to introduce a detailed code of drawings and symbols. As to the use of abbreviations, it is generally suggested that long words (more than 4 to 5 letters according to Rozan (1956/2002, p. 16)) should be abbreviated to save time and effort spent on writing the notes.

The choice of form has also been empirically investigated in such studies as Andres (2002), Dam (Dam, 2004a, 2004b), Lung (2003), Dai and Xu (2007), Liu (2010), and Wang, Zhou, and Wang (2010). The results pointed to a preference for language over symbol, whereas findings on the choice between abbreviation and full word were inconsistent. Most studies recruited student interpreters and some interviewed them afterwards, revealing some potential causes for the preference. Students tended to write down everything as it was heard and were creating symbols on the spot instead of using pre-established symbol systems. Both of these practices limited the use of symbols in note-taking. However, it is questionable whether these findings could be generalised to professional interpreters.

The choice of language is perhaps the most controversial variable in note-taking literature. Traditionally, the categories used to discuss this choice are source and target language. Source language is suggested in some prescriptive literature (e.g., Alexieva, 1994; Gile, 2009; Kirchhoff, 1979) based on the belief that interpreters can "minimize their effort and save capacity" (Szabó, 2006, p. 131) during the listening phase under great time pressure. However, target language is recommended in others (e.g., Herbert, 1952; Jones, 1998; Rozan, 1956/2002) because the authors believe it makes the target speech production phase less effortful, and facilitates better processing of the source speech.

With further empirical data available, some researchers begin to find that the language choice is also affected by whether a language is the A or B language in an interpreter's language combination. In this study, A language refers to the native language while B language refers to the active foreign language. But in order to study the A/B language choice while accounting for the influence of the source/target language status, both directions of interpreting need to be considered, and that has been achieved in only a few studies (e.g., Dam, 2004b; Szabó, 2006; Wang et al., 2010).

Dam (2004b) studied the notes taken by four students with the language combination of Danish/Spanish (three students were Danish native speakers and one was a Spanish native speaker). All her participants preferred the A language regardless of the direction of interpreting, pointing to a tendency to choose the better-mastered native language. Szabó (2006) had eight "quasi professionals" (p. 133) interpret between Hungarian (A language) and English (B language), and all the participants showed a preference for English, their B language, regardless of the direction of interpreting. According to the questionnaire results, participants preferred English because it was "morphologically less complex" and "more economical" (p. 142) than Hungarian, indicating that the nature of the languages themselves played an important role in interpreters' language choice. Wang et al. (2010) studied student interpreters with a language combination of Chinese (A language) and English (B language). They found a source language dominance regardless of the direction of interpreting, and inferred that this could have resulted from the participants' inadequate interpreting competence (p. 15).

The relationship between note-taking and interpreting performance is a key concern in the teaching of interpreting. Scholars have looked at the relationship between interpreting performance and such variables as the quality (Her, 2001) and quantity (Cardoen, 2013; Dam, 2007; Dam, Engberg, & Schjoldager, 2005) of notes, but no consistent conclusions have been reached. It would seem that the interactions between note-taking and interpreting performance are more complex than imagined. A pilot study by Orlando (2014) compared the performances of interpreters in traditional consecutive interpreting and a new hybrid mode using digital pen. Results showed that in the new mode, which he called "consec-simul with notes" (p. 41), the accuracy was higher, and the number of disfluencies or hesitation phenomena was lesser. The digital pen technology was, as a result, recommended for use in consecutive interpreting training and practice.

Through this brief review of literature on note-taking in CI, it is not difficult to find that although some general trends could be detected, such as a dominance of language over symbol, there are also vast inconsistencies. The collected empirical evidence is very limited to date. Many studies that are based on empirical data either use students as participants (whose interpreting competence varies greatly), making the data "not enough to generalise" (Gile, 2009, p. 179), or experiment on one interpreting direction only, making the results difficult to compare.

More importantly, the studies are largely product-oriented. That is, they only look at the product (i.e., the notes produced) without an in-depth analysis of the note-taking process. An outstanding exception was Andres (2002), who used time-coded video to analyse the time span between the moment a source speech unit was spoken (start of sound) and the moment it was noted down (start of pen). She found that, when interpreting from French (B language) into German (A language), the span was between 3 and 6 seconds, although on some occasions it reached as much as 10 seconds. The method used by Andres, however, was to determine the start of note-taking by manually checking a video recording, and the span was measured in seconds, leaving some questions regarding the accuracy of the data.

What could then be a promising avenue for future research? Interpreting is deemed a cognitively demanding task by many. As Gile (2009, p. 178) points out, “note-taking is an area in which the concept of processing capacity can be useful.” If cognitive load can be measured during the process of note-taking, some underlying principles might be unveiled. Considering that discussions on measuring cognitive load in interpreting, especially CI, are very limited (see Chen (2017) for a review and a proposal for potential measurement techniques including pen recording), investigating the cognitive load in note-taking seems important.

This study attempts to address some of the limitations in previous research by (1) using professional interpreters as participants; (2) investigating both directions of interpreting; (3) combining product analysis with the process research method of pen recording; and (4) investigating the cognitive load in note-taking. There are four research questions (RQs), of which the first three are concerned with the three main variables investigated in literature. The aim is to present further empirical data and to either confirm or challenge the previous findings. The fourth RQ pertains to what additional information pen recording can contribute to the topic. The pen data are viewed from a cognitive perspective, and the possibility of using the data as indicators of cognitive load in note-taking and CI is investigated.

RQ1: What do interpreters prefer when choosing the form of note-taking: language or symbol; abbreviation or full word?

RQ2: What do interpreters prefer when choosing the language of note-taking: source or target language; A or B language?

RQ3: What is the relationship between note-taking and interpreting performance?

RQ4: Is there a relationship between the note-taking choices and cognitive load in CI?

## 4.3 Method

As has been mentioned above, in order to make the data more generalizable, research needs to be carried out on professional interpreters (preferably certified and experienced) rather than student interpreters (whose interpreting competence is not yet mature). In order to account for both the source/target language status, and the A/B language status, both directions of interpreting need to be involved. In addition, the note-taking process needs to be recorded. This study was carefully designed to meet those demands.

### 4.3.1 Participants

In this exploratory study, five participants were recruited. They were all certified as "Professional Interpreter" by Australia's National Accreditation Authority for Translators and Interpreters (NAATI). Their working language combination is Mandarin Chinese (A language) and English (B language). Four of them had a postgraduate interpreting degree, and one attended an intensive interpreting training course and obtained a bachelor's degree majoring in interpreting. The participants, aged between 25 and 36 (average 30.2), had worked as full-time or part-time interpreters for three to seven years (average 5.4 years). The city they most frequently worked in was Sydney, Australia. For those who were working as part-time interpreters, their other job(s) involved regular use of both of their working languages (e.g., interpreter trainer). An estimated number of occasions they had provided CI services in the past 12 months ranged from 10 to 50 (average 29).

### 4.3.2 Apparatus

A digital pen and a tablet were used to record pen activities during note-taking. The tablet used was the *Cintiq 13HD* produced by *Wacom*, and it was equipped with a *Wacom Pro Pen*. It was a professional digital tablet targeting graphic designers, developed to meet very high requirements on the precise control of pen strokes. The system has an ergonomic design, with 2048 levels of pressure sensitivity and tilt recognition, closely simulating natural writing and painting.

The *Eye and Pen* software was used to control the whole experiment procedure, and to collect and analyse pen data. The experiment was programmed into the software, which then controlled the procedures to avoid human error. The software can report, for each pen stroke, when the pen tip touches the tablet surface, how it travels across the tablet (distance and duration), and when it leaves the tablet. The spatial data are reported in centimetres and the temporal data are reported in milliseconds. The note-taking and interpreting process was also video-recorded. An additional audio recorder was used to record the retrospective verbal reports (see Section 3.4).

### 4.3.3 Tasks

There were two CI tasks. Stimuli consisted of one Chinese and one English speech, both of which were carefully created through a series of procedures to control for variance.

Firstly, two English video clips on similar topics were selected from the Internet and transcribed by the author. The transcripts were then edited by an experienced university lecturer (a native English speaker from Australia) with respect to length, complexity and style of language, making them as comparable as possible. The edited texts were analysed using *CPIDR*, a computer programme that could automatically determine the propositional idea density, and the results showed that they were quite similar in the number of propositions and words, as well as idea density (Table 4.1).

Table 4.1 Text analysis results

| | **Proposition count** | **Word count** | **Idea density** |
|---|---|---|---|
| **Text 1** | 324 | 630 | 0.514 |
| **Text 2** | 321 | 631 | 0.509 |

Secondly, one of the texts (text 1) was translated by the author into her A language (Chinese), and refined stylistically and grammatically by two Chinese-speaking editors working at a local Chinese radio station. The editors were asked to make the script oral and suitable for recording. They understood the requirements very well due to the nature of their work (editing scripts for radio broadcasting).

Thirdly, the edited Chinese and English scripts were recorded into audio by a native Mandarin Chinese speaker (a radio personality from the same radio station) and a native Australian English speaker (the English editor) in professionally soundproofed studios. The speakers were required to record the speeches as naturally as possible, while maintaining steady speed. They were allowed to restart any sentence at any time when necessary.

Fourthly, the recorded speeches were imported into *Audacity*, a sound-editing programme, for further refinement (e.g., cutting unfinished sentences, deleting background noises). The speeches were both about five minutes long, each divided into three segments (Table 4.2).

Table 4.2 A summary of the tasks

| Task | Topic | Length | Segment length | | |
|---|---|---|---|---|---|
| | | | 1 | 2 | 3 |
| Chinese to English | How to purchase property in Australia | 4m47s | 1m10s | 2m07s | 1m30s |
| English to Chinese | How to register a business in Australia | 4m59s | 1m18s | 2m02s | 1m39s |

### 4.3.4 Procedures

The experiment consisted of three sessions:

Session I: practice. First, the participants were allowed sufficient time to write freely on the tablet using the digital pen. Then, they listened to a short practice task, took notes, and interpreted. The purpose of this step was to get the participants familiarised with both the equipment and the experiment procedures.

Session II: interpreting. The participants first interpreted from Chinese to English. They were allowed a short break if required, and then performed the second task from English to Chinese.

Session III: cued retrospection. Immediately after the tasks, the participants were provided with their notes for cued retrospection. They were asked to provide as much information as they could remember about the note-taking process, including but not limited to: what each note unit was; what it stood for; whether it was symbol or language, and if language, whether it was abbreviation or full word, Chinese or English. This is an important step because note-taking in CI is highly individualised, and the handwriting of interpreters could sometimes be difficult for others to decipher.

### 4.3.5 Data and analysis

The data collected in this study are summarised in Table 4.3. The written notes were analysed to reveal the interpreters' choices of form and language. The distance, duration and speed of pen, and the ear-pen span were used as indicators of the physical, temporal, and cognitive demands of different note-taking choices. Both the notes and the interpreting performance were evaluated by human raters, and analysed together with the note-taking choice results. The qualitative data from retrospection[4] provide an emic perspective from the interpreters, enabling finer-grained analyses of the quantitative data, and help to explain the observed results.

Table 4.3 Data used for analysis

| **Source** | **Data** |
|---|---|
| Pen recording | All written note units;<br>The distance, duration, and speed of each pen stroke;<br>Ear-pen span |
| Video recording | Video of the interpreting process;<br>Audio of the target speech (the interpreting performance) |
| Retrospection | Audio of verbal report |
| Human evaluation | Score of notes;<br>Score of interpreting performance |

#### 4.3.5.1 Categorisation of note units

Based on the interpreters' retrospection, all written notes were categorised according to their form and language (Figure 4.1). Each note unit was first put into one of the three form categories: symbol, language and number. All language note units were further categorised according to form as either abbreviation or full word, and according to language as either Chinese or English.

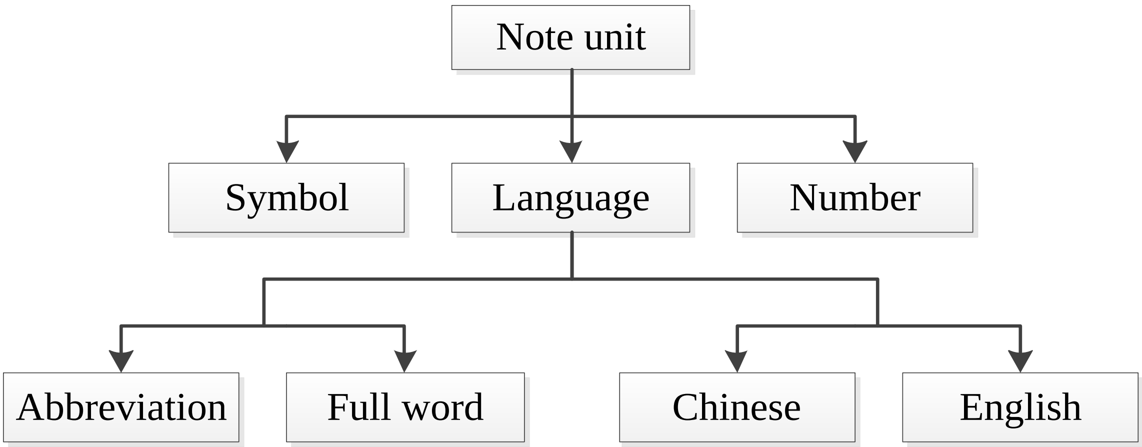


Figure 4.1 Categorisation of note units

[4] The retrospective data in this study is mainly used to assist the researcher to create an accurate interpretation and documentation of the written notes.

The note categories and their definitions are specified in Table 4.4 following the rules specified in Dam (2004a, 2004b). Dam's rules catered to Danish and Spanish, so adaptations were made where necessary to account for the language combination of Chinese and English. For example, Chinese characters with very simple strokes are sometimes used by interpreters as symbols.

Table 4.4 Categories and definitions of note units (Adapted from Dam (2004b, p. 6) and Dam (2004a, p. 253))

| Category | Definition | Examples |
|---|---|---|
| Full word | A full word is a Chinese or English word written in full, including words both with and without morphemes of inflection. | "Problem(s)" and "问题" |
| Abbreviation | An abbreviation consists of parts of the letters of a long English word, or part of the characters of a long Chinese word, or the phonetic spelling of a word, including: (1) real abbreviations (i.e., units in which only part of a word is represented); (2) acronyms; (3) other short forms that cannot be characterised either as real abbreviations or as acronyms, but rather as something in between. | (1) "Prob." / "prblm" for "problem(s)", and "问" for "问题"; (2) "AU" for "Australia", and "澳" for "澳大利亚"; (3) "L&G" for "ladies and gentlemen", and "女&先" for "女士们先生们" ("L" , "G", "女" and "先" will be categorised as abbreviations; "&" will be categorised as a symbol) |
| Symbol | A symbol is a representation of (1) the underlying meaning of a word or expression rather than the actual word or expression; or (2) the relationship(s) between two units. Symbols are mostly pictorial, but they can also be a pair of letters, a single letter, or (part of) a Chinese character. | (1) Signs like pluses and colons, lines, arrows, drawings, etc.; (2) Letter "B" for "but", "however", "on the other hand", "although", etc.; (3) Chinese character "心" for "爱 (love)", "喜欢(like)", "想要 (wanting)", "满意(satisfied)", etc. |
| Language | The combination of full words and abbreviations. Further divided into Chinese and English[5]. | |
| Number | Independent from language and symbol, numbers are seen as a special category of notes. | |

#### 4.3.5.2 Calculation of the ear-pen span

The ear-pen span is defined as the time span between the moment a speech unit is heard (end of sound) and the moment it is written down in notes (start of pen). It was

[5] Unlike in Dam (2004a, 2004b), the author found no notes written in a third language or an unidentifiable language.

calculated using the following steps. First, identifying correspondence between the source speech and the notes. The content and meaning of each note unit (identified with the help from retrospective reports) were checked to determine if there was a one-to-one correspondence between the note unit and a source speech unit. The ear-pen span could not be calculated for notes that did not correspond to specific source speech units (e.g., symbols indicating hidden links).

Second, determining the end of sound and the start of pen. For each note unit that corresponded to a source speech unit, two points in time were determined: (1) the end of sound of the source speech unit; and (2) the start of pen stroke. This was different from what Andres (2002) did in her study, where the time lag was calculated from the start of sound to the start of pen. The consideration was that a span calculated from the start of sound would be heavily influenced by the length of the sound unit. To avoid that influence, this study calculated the span from the end of sound to the start of pen. The start of pen in time was automatically reported by the software in milliseconds. The end of sound was determined by checking the sound waves of the source speech audio using *Audacity*, also reported in milliseconds. The software kept an experiment log which recorded the time that the source speech started to play, so for each note unit, the end of sound and the start of pen could be pinpointed on the same timeline.

Third, calculating the span. The ear-pen span was calculated as "start of pen minus end of sound". It was usually positive, indicating that there was a lag between hearing a source speech unit and noting it down. But on some rare occasions, the span was negative, indicating that the interpreter started to write down the note before hearing the end of a source speech unit, or even predicted an entire incoming unit.

#### 4.3.5.3 Human evaluation

Both the notes and the interpreting performance were rated by two raters: the author and a colleague. Both raters had previous experience of rating interpreter certification tests.

*Rating the notes.* Each note unit that has a one-to-one correspondence with the source speech was rated. It was given a score of either 1 or 0. When a note successfully represented a source speech unit and was correctly interpreted in the target speech, it was scored 1. If it falsely represented the source speech, did not appear in the target speech or was falsely interpreted in the target speech, it was scored 0. For example, if a note unit was written as "invest" (standing for "investment"), and interpreted as "investor" (because the interpreter could no longer identify which meaning it stood for), it would be scored 0.

For each note unit raters were given the content, meaning, corresponding source text unit, source text sentence, and target text sentence (both orthographic transcription). The scores of all note units were added up, and divided by the total number of notes being rated, thus forming the score of notes (i.e., percentage of notes correctly interpreted). The scores given by two raters were averaged.

*Rating the interpreting performance.* The purpose of performance rating in this study was quite different from those in interpreter education or testing. There was no need to judge whether the performance reached certain standards, because all participants were nationally accredited, experienced interpreters. The goal was to differentiate the performances as finely as possible, so that the relationship between note-taking and interpreting performance could be revealed. Considering that all the participants were expected to give high-quality performance, it would be very difficult to use holistic scores to differentiate the performances. A stringent rating system therefore needed to be developed.

The criterion chosen for performance rating in this study was accuracy, a core component of interpreting quality (e.g., Gile, 1999; Pöchhacker, 2002). Many researchers have applied it as a yardstick to evaluate interpreting performance (e.g., Dam & Engberg, 2006; Gerver, 1969/2002; M. Liu & Chiu, 2009). It is also "particularly relevant and central" to studies on note-taking because notes function as memory triggers to ensure an accurate rendition (Dam & Engberg, 2006, p. 216).

The method used for performance rating in this study was a proposition-based one. As has been mentioned, the two original English texts were analysed using *CPIDR*, and the proposition count of the two were 339 and 321 respectively. Based on the proposition analysis results, the texts were divided into scoring units. The rule was that each unit contained an average of three propositions, and natural sentence breaks were kept. The two raters first divided the units separately, and then discussed the inconsistencies and reached agreement. The Chinese text (an edited translation of one of the English texts) was divided following the units marked on the original English text, and the raters discussed the units on the Chinese text and reached agreement. The final number of scoring units were 101 in the Chinese to English task, and 112 in the English to Chinese task.

The interpreting performances were transcribed orthographically by the author and the target texts were provided to the raters for rating. The accuracy was determined by checking how closely each scoring unit was matched by the target text. A score of 1 was given when the meaning of a unit was correctly interpreted; otherwise a score of 0 was given. Following the principles in M. Liu and Chiu (2009), added information was not penalised and erroneous renderings of the same proposition were penalized only the first time they appeared. A performance score was calculated as the percentage of scoring units correctly interpreted. The two raters did a trial rating session individually on some randomly chosen target texts (covering both tasks), discussed the inconsistencies, and reached agreement. The raters then performed all ratings independently. The final score was an average of the scores given by the two raters.

## 4.4 Results and discussion of note-taking choices and their relationship with interpreting performance

This part reports findings that are directly related to previous literature, and answers the first three RQs: interpreters' preferred choice of form in note-taking; interpreters' preferred choice of language in note-taking; the relationship between note-taking and interpreting performance.

### 4.4.1 Choice of form and language

#### 4.4.1.1 Choice between language and symbol

Descriptive statistics concerning the choice between language and symbol of each participant in the two tasks are summarised in Table 4.5. The quantity of notes taken by each participant varied, with participant 5 (P5) taking down the least (120 in task 1 and 112 in task 2), and P2 taking the most (233 in task 1 and 261 in task 2), indicating that note-taking is a highly individualised activity. But the quantity of notes taken in the two tasks was quite similar, both averaged across individuals (177 vs. 179) and within each individual, indicating that the information density of the two tasks are well controlled.

Table 4.5 Distribution over form: language vs. symbol

| | | **Task 1: Chinese to English** | | |
|---|---|---|---|---|
| Participant | Total | Language | Symbol | Number |
| 1 | 197 | 96 (49%) | 88 (45%) | 13 (7%) |
| 2 | 233 | 118 (51%) | 99 (42%) | 16 (7%) |
| 3 | 164 | 79 (48%) | 71 (43%) | 14 (9%) |
| 4 | 172 | 108 (63%) | 52 (30%) | 12 (7%) |
| 5 | 120 | 79 (66%) | 30 (25%) | 11 (9%) |
| Avg. of Task 1 | 177 | 96 (54%) | 68 (38%) | 13 (8%) |
| | | **Task 2: English to Chinese** | | |
| Participant | Total | Language | Symbol | Number |
| 1 | 188 | 113 (60%) | 67 (36%) | 8 (4%) |
| 2 | 261 | 135 (52%) | 118 (45%) | 8 (3%) |
| 3 | 164 | 88 (54%) | 65 (40%) | 11 (7%) |
| 4 | 171 | 119 (70%) | 45 (26%) | 7 (4%) |
| 5 | 112 | 70 (66%) | 30 (27%) | 11 (7%) |
| Avg. of Task 2 | 179 | 106 (59%) | 65 (36%) | 8 (5%) |
| Avg. across participants & tasks | 178 | 100 (57%) | 67 (37%) | 11 (6%) |

Note: percentages may not add up to 100% due to rounding.

As can be seen from the table, there was a clear preference for language over symbol (57% vs. 37% when averaged across participants and tasks), and the trend was consistently reflected in all individual cases and in both directions of interpreting.

#### 4.4.1.2 Choice between abbreviation and full word

There was a preference for abbreviation (34%) to full word (22%) averaged across participants and tasks (Table 4.6). This trend was consistently reflected in both directions, but not in all cases (the exceptions are P3 in task 2 and P5 in both tasks).

Table 4.6 Distribution over form: abbreviation vs. full word

| **Task 1: Chinese to English** | | |
|---|---|---|
| Participant | Abbreviation | Full word |
| 1 | 61 (31%) | 35 (18%) |
| 2 | 89 (38%) | 29 (12%) |
| 3 | 45 (27%) | 34 (21%) |
| 4 | 69 (40%) | 39 (23%) |
| 5 | 37 (31%) | 42 (35%) |
| Avg. of Task 1 | 60 (34%) | 36 (20%) |
| **Task 2: English to Chinese** | | |
| Participant | Abbreviation | Full word |
| 1 | 69 (37%) | 44 (23%) |
| 2 | 95 (36%) | 40 (15%) |
| 3 | 39 (24%) | 49 (30%) |
| 4 | 70 (41%) | 49 (29%) |
| 5 | 37 (33%) | 37 (33%) |
| Avg. of Task 2 | 62 (35%) | 44 (24%) |
| Avg. across participants & tasks | 61 (34%) | 40 (22%) |

Note: percentages do not add up to 100% because the rest are symbols and numbers.

#### 4.4.1.3 Choice of language

The distribution of notes over language (Table 4.7) shows that the participants as a group preferred B language (English, accounting for 36%) to A language (Chinese, accounting for 20%). This preference was consistent in both tasks (i.e., both interpreting directions), and in most participants (with the only exception of P4 in task 1). The trend was stronger in task 2 (18% Chinese vs. 41% English) than in task 1 (23% Chinese vs. 31% English), showing that when the source language and B language coincided, the preference for B language was strengthened.

Table 4.7 Distribution over language

| | **Task 1: Chinese to English** | |
|---|---|---|
| Participant | Chinese | English |
| 1 | 44 (22%) | 52 (26%) |
| 2 | 50 (21%) | 68 (29%) |
| 3 | 32 (20%) | 47 (29%) |
| 4 | 73 (42%) | 35 (20%) |
| 5 | 5 (4%) | 74 (62%) |
| Avg. of Task 1 | 41 (23%) | 55 (31%) |
| | **Task 2: English to Chinese** | |
| Participant | Chinese | English |
| 1 | 48 (26%) | 65 (35%) |
| 2 | 41 (16%) | 94 (36%) |
| 3 | 9 (5%) | 79 (48%) |
| 4 | 42 (25%) | 77 (45%) |
| 5 | 18 (16%) | 56 (50%) |
| Avg. of Task 2 | 32 (18%) | 74 (41%) |
| Avg. across participants & tasks | 36 (20%) | 65 (36%) |

Note: percentages do not add up to 100% because the rest are symbols and numbers.

### 4.4.2 Relationship between note-taking and interpreting performance

As we can see in Table 4.8, neither the score of notes nor the quantity of notes alone could explain the variances in performance:

Table 4.8 Relationship between note-taking and interpreting performance

| Task 1: Chinese to English | | | |
|---|---|---|---|
| Participant | Score of performance | Score of notes | Quantity of notes |
| 1 | 79.21 | 87.27 | 197 |
| 2 | 89.61 | 92.29 | 233 |
| 3 | 81.19 | 87.80 | 164 |
| 4 | 67.83 | 85.57 | 172 |
| 5 | 74.76 | 89.82 | 120 |
| **Task 2: English to Chinese** | | | |
| Participant | Score of performance | Score of notes | Quantity of notes |
| 1 | 87.95 | 95.00 | 188 |
| 2 | 92.86 | 94.48 | 261 |
| 3 | 68.30 | 83.93 | 164 |
| 4 | 69.20 | 88.41 | 171 |
| 5 | 69.65 | 92.22 | 112 |

Rather, performance seemed to be a function of both the quality and quantity of notes. For example, P1 had high note counts, but low note scores, and his/her performance ranked in the middle. P5 had high note scores, but influenced by his/her low note counts, his/her performance also ranked in the middle. P2 had both high note counts and high note scores, and his/her performance ranked the highest.

### 4.4.3 Discussion

Some tentative answers could be suggested for the first three RQs. It needs to be noted that the answers are based on empirical results found on a small sample of professional interpreters working between the language combination of Chinese and English.

RQ1: What do interpreters prefer when choosing the form of note-taking: language or symbol; abbreviation or full word?

Interpreters in our study showed a clear preference for language over symbol. This finding corroborates previous studies, using either student interpreters (Dai & Xu, 2007; Dam, 2004b; J. Liu, 2010; Lung, 2003; Wang et al., 2010) or professional interpreters (Andres, 2002; Dam, 2004a) as participants. The interpreters preferred abbreviation to full word, a finding corroborating some studies (Dai & Xu, 2007; Wang et al., 2010), but contradicting the findings of others (Dam, 2004a; J. Liu, 2010; Lung, 2003). The contradiction could be caused by such factors as the nature of the language pair, the type of participants used, or the texture or genre of the source speech (Setton &

Dawrant, 2016, p. 211), but there is not enough empirical evidence to pinpoint the cause at the moment.

RQ2: What do interpreters prefer when choosing the language of note-taking: source or target language; A or B language?

The interpreters showed a preference for English (their B language) over Chinese (their A language), and this preference was strengthened when the B language and the source language coincided. That is to say, the interpreters opted for a language that is weaker in their language combination, a choice intuitively implausible. Sifting through the retrospective reports, it was found that in many cases, the interpreters chose English for note-taking because it was easier and faster to write than Chinese characters. What also needs to be noted is that the interpreters in this study are based in Australia, an English-speaking country, and they are likely to have a very strong B language.

The result relating to the choice of language in this study contradicts what Wang et al. (2010) found in student interpreters with the same language combination, where a strong preference for source language was detected regardless of the direction of interpreting. It also contradicts what Dam (2004b) found in students with the Danish/Spanish language combination, where a strong preference for the A language was found, indicating a tendency to choose the better-mastered native language. It is in line with what Szabó (2006) found in professional interpreters with the Hungarian/English language combination. Szabó observed a preference for English, the B language, regardless of the direction of interpreting, and pointed to the morphological complexity of Hungarian and the economy of writing in English as an explanation. Szabó also mentioned that the participants had a very strong B language, as is the case in this study.

Based on the above discussion, some conclusions could be suggested on the choice of language in note-taking. The language choice is a function of the combined influence of a series of factors, including: (1) the nature of the languages themselves (e.g., morphological complexity and economy of writing); (2) the A/B language status; (3) the source/target language status; and (4) interpreter characteristics (e.g., working experience and language competence).

When two languages do not differ too much in morphological complexity or economy of writing (the case in Dam (2004b)), the A/B language status plays a major role in determining the language choice, and interpreters are more likely to use their A language for note-taking. When one language is morphologically simpler or easier and faster to write (the case in Szabó (2006) and this study), this language would be the preferred choice regardless of the A/B language status, especially when the interpreter has a strong B language. When the interpreter lacks experience (the case in Wang et al. (2010)), the language choice is subject mainly to the source/target language status in a task.

The empirical data collected so far are insufficient to identify how the factors interact, and what their respective and combined influences are on the choice of language. These are interesting directions for future research.

RQ3: What is the relationship between note-taking and interpreting performance?

With a small sample size of five, it is difficult to draw any concrete conclusions using the data in this explorative study. However, it would seem that the interpreting performance is subject to variances in both the quality and quantity of notes. The following are tentative explanations. The quality of notes is based on two levels of equivalence (between source speech/notes and between notes/target speech). For all notes to faithfully represent the source speech and be successfully rendered in the target speech, an interpreter needs to allocate enough cognitive capacity to activities such as listening/analysing and memorising. This would sometimes lead to a decrease in the amount of notes that can be written down, reducing the amount of information that can be stored in notes. A good interpretation is related to the concurrent storage of information both in notes and in memory. That is why sometimes we could observe a set of notes with high score but low quantity to be associated with a middle-ranking performance. Previous studies have also detected potential relationships between performance and the quality (Her, 2001) and quantity (Dam, 2007) of notes. But the interactions between the variables and their individual and combined influences on performance remain unclear with the available data. Further empirical evidence needs to be gathered before the mechanism could be revealed.

## 4.5 Results and discussion of the pen data: potential indicators of cognitive load in CI

This part reports on data collected via pen recording. Different note-taking choices are compared on the distance, duration, and speed of pen, as well as the ear-pen span. The data are examined from a cognitive perspective, with an attempt to answer the last RQ:

RQ4: Is there a relationship between the note-taking choices and cognitive load in CI?

### 4.5.1 Pen data on the choice of form

#### 4.5.1.1 Between language, symbols, and numbers

Consistent differences are found between language and symbol notes in terms of the distance, duration, and speed of pen, and the ear-pen span (Table 4.9). The average distance and duration of language notes (7.17 cm and 1256.13 ms respectively) were much longer than those of symbol notes (2.99 cm and 367.48 ms respectively), and the writing speed of symbol (9.14 cm/s) was faster than that of language (6.04 cm/s). That is to say, compared to language, symbols are easier and faster to write. The ear-pen span of symbol (3039.33 ms) was longer than that of language (2504.99 ms), indicating it took longer for interpreters to transfer a source speech unit into symbol than into language notes.

Interestingly, the distance, duration, and speed of pen of numbers all lie between those of language and symbol, but the ear-pen span of numbers (1428.31 ms) was much shorter than both. This means that, after hearing a number, the participants would take very swift responses and write it down, about one second faster than language and 1.5 seconds faster than symbols.

Table 4.9 Pen data on the choice of form

| | Task 1: Chinese to English | | | | |
|---|---|---|---|---|---|
| | **Language** | **Symbol** | **Number** | **Abbreviation** | **Full word** |
| Distance (cm) | 7.17 | 3.21 | 4.97 | 6.48 | 8.14 |
| Duration (ms) | 1237.56 | 379.04 | 796.99 | 1094.46 | 1433.94 |
| Speed (cm/s) | 6.20 | 9.13 | 6.64 | 6.22 | 6.21 |
| Ear-pen span (ms) | 2620.10 | 2980.78 | 1682.21 | 2412.34 | 2925.95 |
| | **Task 2: English to Chinese** | | | | |
| | **Language** | **Symbol** | **Number** | **Abbreviation** | **Full word** |
| Distance (cm) | 7.17 | 2.77 | 5.24 | 6.03 | 8.66 |
| Duration (ms) | 1274.70 | 355.93 | 1021.87 | 1088.51 | 1527.51 |
| Speed (cm/s) | 5.88 | 9.15 | 6.10 | 5.80 | 5.91 |
| Ear-pen span (ms) | 2389.87 | 3097.89 | 1174.41 | 2436.28 | 2356.79 |
| | **Averaged across participants and tasks** | | | | |
| | **Language** | **Symbol** | **Number** | **Abbreviation** | **Full word** |
| Distance (cm) | 7.17 | 2.99 | 5.10 | 6.25 | 8.40 |
| Duration (ms) | 1256.13 | 367.48 | 909.43 | 1091.48 | 1480.72 |
| Speed (cm/s) | 6.04 | 9.14 | 6.37 | 6.01 | 6.06 |
| Ear-pen span (ms) | 2504.99 | 3039.33 | 1428.31 | 2424.31 | 2641.37 |

#### 4.5.1.2 Between abbreviations and full words

Consistent differences are found between abbreviation and full word notes in terms of the distance and duration of pen (Table 4.9). The average distance and duration of pen of abbreviations (6.25 cm and 1091.48 cm respectively) were shorter than those of full words (8.40 cm and 1480.72 cm respectively), indicating that abbreviations were easier to write, but the speed of pen was similar (6.01 cm/s for abbreviations and 6.06 cm/s for full words).

In task 1, the average ear-pen span of abbreviations (2424.31 ms) was shorter than that of full words (2641.37 ms), and this difference was consistent in all participants. However, no consistent trend could be detected in task 2.

### 4.5.2 Pen data on the choice of language

The distance, duration, and speed of pen showed no consistent difference between the language choices. But the ear-pen span (Table 4.10) of notes in A language was longer than B language in almost all cases (except for P1 in task 2). That is to say, after hearing a source speech unit (no matter in what language), it takes longer before the participants write down a Chinese note than an English one.

Table 4.10 Ear-pen span data on the choice of language

| Task 1: Chinese to English | | |
|---|---|---|
| Participant | Chinese | English |
| 1 | 2137.35 | 2105.78 |
| 2 | 2489.00 | 2390.92 |
| 3 | 2774.55 | 2179.00 |
| 4 | 3104.04 | 3055.06 |
| 5 | 3808.80 | 2983.53 |
| Avg. of Task 1 | 2862.75 | 2542.86 |
| **Task 2: English to Chinese** | | |
| Participant | Chinese | English |
| 1 | 1855.02 | 2717.71 |
| 2 | 2245.11 | 2142.57 |
| 3 | 2648.63 | 1584.61 |
| 4 | 2969.07 | 2476.84 |
| 5 | 3227.28 | 3035.06 |
| Avg. of Task 2 | 2589.02 | 2391.36 |
| Avg. across participants & tasks | 2725.88 | 2467.11 |

### 4.5.3 Interpreting the findings from a cognitive load perspective

No matter what choices interpreters make during note-taking, the basic question, as Gile (2009, p. 178) points out, is "how to reduce processing capacity and time requirements of note-taking while maintaining the efficiency of notes as memory reinforcers". On the cognitive side, since interpreting is a highly demanding task, an important goal of interpreters' skills and strategies is to save cognitive effort. On the physical and temporal side, the physical effort and time cost associated with note-taking are of great concern to consecutive interpreters (Alexieva, 1994). Therefore we have good reasons to infer that, for professional interpreters with sufficient experience, their overall choices should reflect a balanced weighting of the physical, temporal and cognitive demands of note-taking.

In this study, the distribution data showed that interpreters preferred language (57%) to symbol (37%), abbreviation (34%) to full word (22%), and English (36%) to Chinese (20%) during note-taking. We would like to make the bold hypothesis that the overall physical, temporal and cognitive demands associated with different note-taking choices for Chinese interpreters working between English and Chinese is: language lower than symbol, abbreviation lower than full word, and English lower than Chinese, regardless of the direction of interpreting.

The pen data of distance and duration could be straightforward indicators of the physical effort and temporal cost associated with the note-taking choices. Notes that induce lower demands should be those with shorter pen distance and duration, meaning the pen tip travels a shorter distance and for a shorter period of time. According to our results, the distance and duration of language and full words are longer than those of symbols and abbreviations respectively, suggesting that the use of symbols and abbreviations could reduce physical and temporal demands. This finding corroborates the note-taking principles proposed by many (e.g., Alexieva, 1994; Gillies, 2005; Schweda-Nicholson, 1993). No clear difference is found in the physical and temporal demands between Chinese and English notes, suggesting that the choice of language does not significantly affect the physical or temporal demand of note-taking.

The ear-pen span data are potentially indicative of the cognitive load in note-taking. Since interpreting is an externally paced task, high cognitive load tends to increase the time lag, causing participants to "lag farther and farther behind the input" (Treisman, 1965, p. 378). In our study, the ear-pen span results were: symbol longer than language, full word longer than abbreviation, and Chinese longer than English. Assuming the ear-pen span is an indicator of cognitive load (longer span means higher load), then the cognitive load associated with different note-taking choices are: language lower than symbol, abbreviation lower than full word, and English lower than Chinese.

If we put the two pieces of the puzzle together (Table 4.11), we can see how the physical, temporal and cognitive demands act together to affect interpreters' note-taking choices. It would seem that physical and temporal demands do not affect note-taking as much as cognitive load. In particular, despite their lower physical and temporal demands, symbols are used less than language by interpreters in note-taking.

Table 4.11 How physical, temporal and cognitive demands affect interpreters' note-taking choices

| | Form | | Language |
|---|---|---|---|
| | Language vs. symbol | Abbreviation vs. full word | Chinese vs. English |
| **Physical and temporal demands** | Symbol < Language | Abbreviation < Full word | Chinese ≈ English |
| **Cognitive load** | Language < Symbol | Abbreviation< Full word | English < Chinese |
| **Note-taking preference** | Language | Abbreviation | English |

## 4.6 Conclusions

This study investigates note-taking in CI in terms of the choice of form and language, and the relationship between note-taking and interpreting performance. It reports new data from pen recording, interprets the data from a cognitive perspective, and presents preliminary findings on the relationship between note-taking and cognitive load.

It was found that, firstly, interpreters preferred language to symbol, abbreviation to full word, and English to Chinese, regardless of the direction of interpreting. Secondly, the interpreting performance seemed to be subject to variances in both the quality and quantity of notes. Thirdly, the physical and temporal demands of different note-taking choices, as indicated by the pen data of distance and duration, appeared to be: language higher than symbol, full word higher than abbreviation, and Chinese similar to English. Fourthly, the cognitive load induced by different note-taking choices, as indicated by the ear-pen span, appeared to be: symbol higher than language, full word higher than abbreviation, and Chinese higher than English.

On the whole, pen recording is found to be a powerful method to tap into the process of note-taking and CI. The data collected can provide us with an accurate and encompassing picture of the interpreting process with moment-to-moment changes in pen position reported in coordinates. The data also appear to be useful indicators of cognitive load. Although the digital pen and tablet system used in this study is particularly useful for research purposes, it is not recommended for application in training or practice. There are other types of digital pens which are less powerful in data collection but much easier to use in classrooms and field interpreting (see Orlando, 2010, 2014).

It has to be admitted that the empirical data collected in this study are very limited. The sample size is small, and only one language combination (Chinese and English) is involved, confining the generalizability of the findings. But at the same time, the limitations have pointed to some interesting directions for future research. For example, can the findings be replicated with a larger sample size? Can the same results be reached using a different language combination? The author will continue to seek answers to these questions, and hopefully they will attract the interests from other researchers as well.

### 4.7 References


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting, 14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Albl-Mikasa, M. (2008). (Non-) sense in note-taking for consecutive interpreting. *Interpreting*, *10*(2), 197–231. doi:10.1075/intp.10.2.03alb

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199–206). Amsterdam: John Benjamins.

Andres, D. (2002). *Konsekutivdolmetschen und notation [Consecutive interpreting and note-taking]*. Frankfurt: Peter Lang.

Cardoen, H. (2013). The Effect of Note-taking on Target-text Fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging Research in Translation Studies: Selected Papers of the CETRA Research Summer School 2012* (pp. 1-22). Leuven: CETRA.

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

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*. doi:10.1080/0907676X.2016.1278026

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research, 39*(2), 136–144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting, 6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters’ notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V., & Engberg, J. (2006). Assessing accuracy in consecutive interpreting: a comparison of semantic network analyses and intuitive assessments. In C. Heine, K. Schubert, & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp. 215–234). Tübingen: Narr Francke Attempto Verlag.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters’ notes. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (Vol. 7, pp. 227–254). Berlin: Mouton de Gruyter.

Gerver, D. (1969/2002). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 53–66). London and New York: Routledge.

Gile, D. (1991). Prise de notes et attention en début d’apprentissage de l’interprétation consécutive–une expérience– démonstration de sensibilisation [Note-taking and attention at the beginning of consecutive interpreting learning –an experience–demonstration of awareness]. *Meta, 36*(2–3), 431–439.

Gile, D. (1999). Variability in the perception of fidelity in simultaneous interpretation. *Hermes, 22,* 51–79. doi:10.7146/hjlcb.v12i22.25493

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

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. Manchester and Kinderhook: St. Jerome.

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation, 6*, 53-77.

Herbert, J. (1952). *Manuel de l'interprète: comment on devient interprète de conférences [Interpreter's manual: How to become a conference interpreter]*. Geneva: Georg.

Jones, R. (1998). *Conference interpreting explained*. Manchester: St. Jerome.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Kohn, K., & Albl-Mikasa, M. (2002). Note-taking in consecutive interpreting: On the reconstruction of an individualised language. *Linguistica Antverpiensia, 1*, 257-272.

Liu, J. (2010). 英语专业本科学生汉英交传笔记特征——一项基于学生交传笔记的实证研究 [Note-taking characteristics of English majored undergraduates in Chinese-English consecutive interpreting: an empirical study based on students' consecutive interpreting notes]. *Foreign Language World, 2*, 47–53.

Liu, M., & Chiu, Y.-H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting, 11*(2), 244–266. doi:10.1075/intp.11.2.07liu

Lung, R. (2003). Taking "notes" seriously in the interpretation classroom. In Á. Collados Aís, M. M. Fernández Sánchez, & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación [The evaluation of quality in interpretation: Research]* (pp. 199–205). Granada: Comares.

Matyssek, H. (1989). *Handbuch der notizentechnik für dolmetscher [Handbook of note-taking for interpreters]*. Heidelberg: Julius Groos.

Orlando, M. (2010). Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter, 15*, 71–86.

Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation and Interpreting, 6*(2), 39–54.

Pöchhacker, F. (2002). Researching interpreting quality: Models and methods. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 95–106). Amsterdam: John Benjamins.

Rozan, J.-F. o. (1956/2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Cracow: Tertium Society for the Promotion of Language Studies.

Schweda-Nicholson, N. (1993). An introduction to basic note-taking skills for consecutive interpretation. In E. Losa (Ed.), *Keystones of communication: Proceedings of the 34th Annual Conference of the American Translators Association* (pp. 197–204). Medford: Learned Information.

Seleskovitch, D. (1975). *Langage, langues et mémoire: étude de la prise de notes en interprétation consécutive [Speech, language, and memory: A study of note-taking in consecutive interpreting]*. Paris: Minard Lettres Modernes.

Setton, R., & Dawrant, A. (2016). *Conference interpreting: A trainer's guide*. Amsterdam: John Benjamins.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting, 8*(2), 129–147. doi:10.1075/intp.8.2.02sza

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology*, *56*(4), 369–379. doi:10.1111/j.2044-8295.1965.tb00979.x

Wang, W., Zhou, D., & Wang, L. (2010). 口译笔记特征与口译产出质量实证研究 [An empirical study of note-taking characteristics and output quality in interpreting]. *Foreign Language World, 4*, 9–18.

## An introductory note to Chapter 5

Findings in Chapter 4 allow specific hypotheses to be realised about the form and language of note-taking and its relationship with interpreting performance. With these hypotheses available, the PhD research moves on to the next level to carry out a main study involving a larger sample of participants. The main study implemented the full-fledged design of the PhD project, triangulating the methods of pen recording, eye tracking and voice recording to explore the cognitive processing and cognitive load in note-taking and CI.

Chapter 5 reports the pen and voice recording data of the main study, focusing on Phase I of CI in which interpreters listen to the source speech and write notes. Issues investigated include the content of notes, the timing of note-writing, and the choices of form and language in note-taking as well as their associated cognitive load. Most of the findings in the pilot study (Chapter 4) are successfully replicated in the main study. Several trends in the seemingly individualised note-taking choices are also unveiled. Discussions on how the findings in this study add to the literature on note-taking and CI and pedagogical suggestions are provided on the basis of the empirical data.

# Chapter 5 The process of note-taking and consecutive interpreting: Evidence from digital pen recording[1]

**Abstract:** This article reports the findings of an empirical study on the process of note-taking and consecutive interpreting (CI). Combining digital pen recording with video recording, the note-taking and interpreting process is recorded in detail and automatically synchronised with the source speech. The study seeks answers to four questions about CI and note-taking: (1) What do interpreters note down? (2) When do interpreters take notes? (3) How do interpreters take notes? (4) Why do interpreters make certain choices in note-taking? Data that have been collected include the note-taking and interpreting process (e.g., all note units; details of the pen movement which reveal information on the physical, temporal and cognitive demands of different note-taking choices) and the interpreting performance. By triangulating the different types of data, some general trends in the process of note-taking and CI were detected, and some fundamental principles guiding interpreters' note-taking choices were unveiled. This forms the basis for a number of pedagogical recommendations.
**Keywords:** note-taking; consecutive interpreting; process research; cognitive load; digital pen recording

## 5.1 Introduction

The field of translation and interpreting process research has seen significant development over the past years and in particular, the focus on the cognitive aspects of translation and interpreting is increasing. This is evident in the growing number of volumes that have been published in the last two decades (e.g., Ehrensberger-Dow, Dimitrova, Hubscher-Davidson, & Norberg, 2015; Martín, 2016; Shreve & Angelone, 2010; Tirkkonen-Condit & Jääskeläinen, 2000). Researchers have benefited from an interdisciplinary effort, which resulted in the application of many methods that are new to the field, such as verbal reporting, key logging, video and screen recording and eye tracking.

However, in the field of interpreting studies, consecutive interpreting (CI) has received very limited attention so far. To look into the process of CI, note-taking serves

[1] This chapter is under review in the journal of *Interpreting* as: Chen, S. (under review). The process of note-taking and consecutive interpreting: Evidence from digital pen recording.

as a good starting point. It is a unique characteristic of CI[2] and can reveal important information about the cognitive processes. Consecutive note-taking has been a recurring topic in the field, creating fruitful results over the past half a century. But most of the studies have been product-oriented, without collecting and analysing process data. A potentially important reason for this is a lack of existing methods to tap into the process of CI within the field of interpreting research. Nevertheless, there are a handful of studies that have taken a process-oriented approach to note-taking and CI. Two outstanding examples would be Andres (2002) who used video recording and Orlando (2010) who used the *Livescribe Smartpen*. However, neither method could provide comprehensive data on the entire process. For example, video recording involves determining the start of note-taking by manually checking the video and its timestamp and the *Smartpen* does not report the moment-to-moment change in pen position in coordinates.

In writing research, there is a new pen recording method which involves the use of a software called the Eye and Pen[3]. With the software, a digital tablet and a digital pen, any writing process can be recorded in fine details (with time data reported in milliseconds and position data reported in centimetres). This study applies this method to investigate the process of note-taking and CI. The software is programmed to control the play of the source speech (hence listening) and align it with the note-taking process (for more details please see Section 3.2). With this synchronous recording of the multiple tasks happening in the interpreting process, this study explores the what, when and how questions about CI and note-taking, and tries to find an explanation to why interpreters make certain note-taking choices.

## 5.2 Research background

### 5.2.1 A review of studies on note-taking in CI

There are two main streams of note-taking literature: a prescriptive stream and a descriptive stream (for a detailed review please see Chen (2016)). In the prescriptive stream a number of books and articles put forward various note-taking systems and principles. Prescriptive suggestions are given on what, when and how to take notes, usually starting from the authors' experiences in the profession and/or in teaching.

[2] In this article, CI refers to long consecutive in which systematic note-taking is used.

[3] More introduction to the software can be found on http://www.eyeandpen.net/?lng=en.

The first and foremost principle in any note-taking system is to take notes based on speech analysis. This is described by Gillies (2005) as identifying the "Subject Verb Object" unit. Things that are recommended to be noted down are (1) those to cue the interpreter and ensure an efficient rendering of the speech, including the main ideas, links, verb tenses and modal verbs; and (2) those to relieve memory, including numbers, dates, proper names and lists (e.g., Gillies, 2005; Jones, 1998; Rozan, 2002).

Slightly different suggestions have been given on when to take notes. Some recommend starting to take notes as quickly as possible (Jones, 1998, pp. 61-63); others suggest waiting until the subject of a speech is fully understood (Lung, 1999, p. 311); still others suggest that notes can be taken either sooner or later according to the circumstances (Gillies, 2005, pp. 156-158). It is generally agreed that crucial details (especially those difficult to store in memory) such as names, numbers and dates should be noted down as soon as possible and that interpreters should not be bound by the original order of things in the source speech. They are free to change the order to coordinate note-taking and memory (e.g., Gillies, 2005, p. 159; Jones, 1998, p. 63).

Discussions on how to take notes are generally focused on the use of symbols and abbreviations, as well as the language of note-taking. The use of symbols is an essential part of any established note-taking system. Symbols are easy to write and read, and because they represent concepts rather than words, they can help to avoid source language influence (Gillies, 2005, p. 99). Although the amount of exemplary symbols given by different authors varies (for example Rozan (2002) on the minimalist end and Matyssek (1989) on the maximalist end), there is general agreement that fully-mastered, unambiguous symbols are very useful. The use of abbreviations is suggested for noting long words. This could be done by writing the first and last letters, preferably with the latter written as superscript (Matyssek, 1989, p. 115; Rozan, 2002, pp. 16-18; Schweda-Nicholson, 1993, p. 200), and by using phonetic spelling and misspelling (Gillies, 2005, p. 162). As to the language of note-taking, authors tend to not give any specific suggestions, although sometimes a slight preference is given to either the source language (e.g., Alexieva, 1994; Ilg, 1988) or the target language (e.g., Rozan, 2002; Seleskovitch, 1975; Seleskovitch & Lederer, 1989/1995).

It would seem that the note-taking systems are well-developed, and once students are made aware of them and practice accordingly, note-taking should not be a problem. However, when it comes to the teaching and learning of note-taking, both the teachers and the students find it challenging. The problem has been documented in a number of studies (e.g., Alexieva, 1994; Gile, 1991), in which researchers found that note-taking could divert the attention of student interpreters and could even lead to a degradation in interpreting performance. Studies that describe how notes are actually taken by student interpreters represent the beginning of a shift from prescriptive to descriptive approaches in note-taking literature. This shift is strengthened by researchers who approach the topic from cognitive and linguistic perspectives.

Motivated by an attempt to theorise note-taking and CI, some researchers started investigations on the cognitive and linguistic aspects of note-taking. Studies by pioneering scholars such as Kirchhoff (1979) and Seleskovitch (1975) and the more recent ones by Albl-Mikasa (2006, 2008) have pointed to a concurrent storage of information in notation texts and in memory, as well as a competition for resources between note-taking and other activities in the interpreting process.

Unlike the early descriptive studies which have a general interest in what real notes look like, and set out to discover some overall trends, the latest studies on note-taking usually have more specific targets. They investigate different note-taking choices, collect data in simulated interpreting tasks, and have contributed some valuable empirical evidence. In this group of studies, variables that have attracted the most attention include the choice of form, the choice of language, and the relationship between note-taking and interpreting performance.

The choice of form refers to the choice between language and symbol and the choice between abbreviation and full word. Overall, studies found a dominance of language over symbol but reached inconsistent conclusions as to whether interpreters preferred abbreviation or full word (Chen, 2016).

The choice of language refers to the choice between source and target language and the choice between L1 and L2[4]. It has been found that the notes of student interpreters are largely source-language dominated whereas the notes of professional interpreters are more varied. While some studies found that professionals preferred L1 (e.g., Abuín González, 2012; Dam, 2004a), others found that they preferred L2 such as English in Szabó (2006).

Research on the relationship between note-taking and interpreting performance appears to be even more inconclusive. Some studies claim to have found evidence

[4] In this article, L1 refers to the native language and L2 refers to the non-native language. They are called A and B language in some studies.

pointing to a potential link between certain note-taking features and the quality of interpreting (Cardoen, 2013; Dam, 2007; Dam, Engberg, & Schjoldager, 2005; Her, 2001); others have failed to detect any such relationship (Dai & Xu, 2007).

Most of these studies are product-oriented, meaning they only look at the notes produced without an in-depth analysis of the interpreting process. An outstanding exception to this was Andres (2002), who used time-coded video to analyse the time lag between listening and note-taking in CI. She found that, when interpreting from French (L2) into German (L1), the lag was between 3 and 6 seconds, although on some occasions it reached as much as 10 seconds. Since the time lag in CI refers to the span between the moment a source speech unit is heard and the moment it is written down, it could be called the ear-pen span (EPS). It is similar to the ear-voice span (EVS) in simultaneous interpreting, which has been studied by many (e.g., Barik, 1973; Christoffels & De Groot, 2004; Gerver, 1969/2002; Goldman-Eisler, 1972). EVS and EPS can provide rich information about the cognitive processing in both simultaneous (Timarová, Dragsted, & Hansen, 2011; Treisman, 1965) and consecutive (Chen, 2017) interpreting. In this study, EPS will be carefully analysed and used as an indicator of cognitive load in CI.

On the whole, the amount of empirical data that has been collected on the topic is still limited. Studies that do collect empirical data are usually limited in several aspects: the participants, the method and technology used, the stimulus task, and the statistical analysis. Many of the studies used students as participants, rendering the data insufficient to allow for generalisation (Gile, 2009, p. 179). Studies that have involved professional interpreters usually have small sample sizes (less than 15 participants). The method and technology used cannot capture the note-taking and CI process with enough details, leading to a lack of multiple strands of empirical evidence to elucidate the same process. The stimulus task often involves one interpreting direction only, making it difficult to compare the results and to consider the variation brought by directionality. More importantly, the specifics of the stimulus task (such as length and speed) are simply not reported in many studies, so it is impossible to replicate the experiment. The recorded empirical data are usually descriptive and only demonstrate some general trends (e.g., a preference for L1 or L2) but on many occasions, no significance testing has been performed. All these limitations have confined our ability to understand CI and note-taking in more depth.

### 5.2.2 Research questions

Considering the limitations in the previous studies, this study is designed to address some of the challenges. It uses professional interpreters as participants and combines the process research method of pen recording with product analysis on both the notes and the interpreting performance. It involves two directions of interpreting and performs significance testing on all the results. With these attempts, the study will try to answer four questions about note-taking and CI.

First, what do interpreters note down? This question looks at the linguistic features of interpreters' notes in relation to the source text. Attempts will be made to analyse which source text elements are noted down and whether the findings conform to what has been suggested in the literature.

Second, when do interpreters take notes? This question examines how interpreters coordinate the parallel tasks in the first phase of CI, namely listening/analysis and note-taking. Special attention will be paid to how far interpreters lag behind the source speech in CI and to what extent they engage in multi-tasking.

Third, how do interpreters take notes? This question is concerned with the choice of form (whether interpreters prefer language or symbol, abbreviation or full word) and the choice of language (whether interpreters prefer the source or target language, L1 or L2) in note-taking.

Fourth, why do interpreters make certain note-taking choices? The physical, temporal and cognitive demands associated with the different note-taking choices will be compared to seek possible explanations for interpreters' note-taking preferences. The interpreting performance will also be examined to see if there is a relation between note-taking choices and the quality of interpreting.

## 5.3 Method

### 5.3.1 Participants

The study involves 26 professional interpreters. They were paid for their participation. They all had NAATI Professional Interpreter accreditation, with a working language combination of Mandarin Chinese (L1) and English (L2). Most of them had a postgraduate degree in interpreting (65%); some had an interpreting diploma (15%); some attended an intensive interpreting training course (15%); and one was self-trained (4%) . With an average age of 36.4, the participants had worked as full-time or part-time interpreters for an average of 7.4 years. The country they most frequently worked in was Australia (with only two exceptions who worked more frequently in China). For those who were working as part-time interpreters, their other job(s) had a bilingual feature (e.g., interpreter trainer). The number of occasions they had provided CI services in the past 12 months was averaged at 167.

The participants used a digital pen and a tablet for note-taking. A post-hoc questionnaire was used to ask whether the interpreters felt comfortable with the digital pen and tablet. If the rating was too low (lower than 4 in a 1-7 scale rating[5]), the participant was excluded from the data analysis. The final pen recording data came from 22 participants.

### 5.3.2 Apparatus

The digital pen used in the study was the *Wacom Pro Pen* and the tablet was the *Cintiq 13HD*. The system is ergonomically designed to mimic natural writing and painting, targeting graphic designers who have very high requirements on the precise control of the pen on the tablet surface. It was linked to a laptop computer piloted by *Eye and Pen* software. The experiment procedure was programmed into the software, which controlled the experiment and interacted with the participants. For example, the software would play a new segment of the source speech when participants indicated that the interpretation was completed (by clicking a button displayed on the screen with the pen). It could also create as many new pages as needed when participants listened to the source speech and took notes. The software was also responsible for reporting the moment-to-moment position and state of the pen tip on the tablet and collecting the data. The interpreting process was video-recorded. An additional audio recorder was used to record the retrospective verbal reports (see Section 3.4).

### 5.3.3 Tasks

A series of procedures were carried out to create two comparable source speeches. First, two English scripts (1 and 2) on similar topics were created and edited by an Australian university lecturer with respect to length, complexity and style of language. The resulting scripts were put into a programme called CPIDR 5 for analysis, and results

[5] The question was: "On a scale from 1 (strongly disagree) to 7 (strongly agree), do you agree that the digital tablet and pen are sufficiently similar to real pen and paper, and therefore did NOT affect your note-taking behaviour?"

showed that they were comparable in word count, proposition count and idea density[6] (Table 5.1). Second, script 2 was translated into Chinese, refined stylistically and grammatically by two Chinese editors working at a local Chinese radio station. Third, an English speech was recorded using script 1 by a native English speaker in Australia (the same person who edited the original English scripts). A Chinese speech was recorded by a native Chinese (Mandarin) speaker (a radio personality from the radio station mentioned above) using the Chinese version of script 2. Both recordings were made in professional sound-proof studios. Fourth, the recordings were edited using *Audacity*. Each speech was divided into three segments and controlled for variables such as pauses, duration and speed (Table 5.1).

Table 5.1 A summary of the two tasks

| Task | Topic | Word/character count | Proposition count | Idea density | Duration | Segment duration | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 |
| E-C | How to register a business in Australia | 631 | 321 | 0.509 | 4'59" | 1'18" | 2'02" | 1'39" |
| C-E | How to purchase a property in Australia | 630 (English) * 944 (Chinese) | 324* | 0.514* | 4'47" | 1'10" | 2'07" | 1'30" |

* The calculations were based on the original English script to enable comparison between the two tasks.

### 5.3.4 Procedures

First, participants were allowed sufficient time to practice using a short task. During this practice session they get familiarised with the digital pen and tablet and the experiment procedures. Second, the participants performed the CI tasks, the order of which was randomised. Third, after both tasks were completed, the participants were provided with their written notes for cued retrospection. During the retrospection, the participants told the experimenter as much as they could remember about the note-taking process, including but not limited to: what each note unit was; what it stood for; whether it was symbol or language, and if language, whether it was abbreviation or full word, Chinese or English. This is an important step because note-taking in CI is highly individualised and interpreters' notes could be difficult for others to identify.

[6] More details about the programme and the analysis can be found at http://ai1.ai.uga.edu/caspr/ and Brown, Snodgrass, Kemper, Herman, and Covington (2008).

### 5.3.5 Data and analysis

Data in this study came from several sources: pen recording, video recording, human evaluation, and verbal report. Pen recording produces all the written note units (see 3.5.1) and details of the pen movement (see 3.5.2 and 3.5.4). The video recording of the interpreting process was used to produce transcripts of the interpreting performance and was provided to a group of raters for evaluation (see 3.5.4). The verbal report produced via cued retrospection provided additional subjective data about the interpreting process and served as a basis for categorising the note units (see 3.5.3).

#### 5.3.5.1 Tagging the source speech

All note units with a one-to-one correspondence with the source speech (identified with the help of retrospective reports) were mapped onto the source speeches, which were analysed using part-of-speech (POS) tagging. The POS tags are those of the Penn Treebank (Santorini, 1990; Xia, 2000). The most important ones in this study are: adjective, adverb, conjunction, determiner, noun, number, preposition, pronoun, proper noun, verb, localiser and measure word (for the Chinese speech only), and modal (for the English speech only). The data can reveal which elements in the source speech have a higher frequency of being noted down and whether the empirical findings conform to the prescriptive literature. For example, to see if interpreters give priority to numbers, each source speech element tagged as "number" is checked to see the percentage of interpreters who have noted it down. This percentage is then averaged across all elements tagged as numbers in the source speech, producing the frequency of numbers to be noted down by all interpreters in the task. If the frequency is very high, it is proof that when deciding what to note, interpreters do give priority to numbers.

#### 5.3.5.2 Calculating the EPS

The source speech and the note-taking process were mapped onto the same timeline. The EPS was calculated as the time span between the moment a speech unit is heard and the moment it is written down in notes. As has been mentioned earlier, Andres (2002) used time-coded video to analyse the time span in CI. She calculated the span from the start of sound to the start of pen, and determined the time points by manually checking the videos. But this raises the concern that the data would be influenced by the length of the source speech unit and that the accuracy and objectivity of the data could be affected by human error. To improve the situation, the measurement points in this study were from the end of sound to the start of pen using data automatically reported by software.

Since the EPS can only be calculated for those note units that have a one-to-one correspondence with the source speech, the number of notes included in EPS analysis is smaller than the total number of notes. Moreover, time lag data is sensitive to extreme values, as has been pointed out by researchers who studied EVS in simultaneous interpreting (Oléron & Nanpon, 1965/2002, p. 47; Timarová et al., 2011, p. 142). In note-taking, extreme values are often observed in cases such as note additions at the end of a source speech segment. Therefore, for the analysis of EPS, all observations that were three standard deviations below or above the mean for each condition (a certain participant in a certain task) were excluded from analysis. In the end, about 90% of the notes were included in the EPS analysis.

#### 5.3.5.3 Categorising the notes

Each written note was first categorised according to form as symbol, language, or number[7]. Then, if a note unit was labelled as language, it was further categorised as either abbreviation or full word (according to form) and as either Chinese or English (according to language). The categories and their definitions followed the rules specified in Dam (2004a, 2004b), and adaptations were made where necessary to account for the language pair of Chinese and English. For example, Chinese characters with a few simple strokes are sometimes used by interpreters as symbols.

#### 5.3.5.4 Detailed pen recording data and the interpreting performance

For each note unit, the experiment software recorded the distance of pen (how far the pen moved across the surface, reported in centimetres) and the duration of pen (for how long the pen moved, reported in milliseconds). The distance and duration of pen were used as indicators of the physical and temporal demands of different note-taking choices. Notes that induce lower demands should be those with shorter pen distance and duration, meaning the pen tip moves a shorter distance and for a shorter period of time. The EPS was used as an indicator of the cognitive load. Since interpreting is an externally paced task, high cognitive load tends to increase the time lag, causing participants to "lag

[7] Numbers are put into an independent category because they are often treated with special care by interpreters, which is evident both in literature (see Section 2) and from the results in this study (see Section 4.2).

farther and farther behind the input" (Treisman, 1965, p. 378). Notes that induce lower cognitive load should be those with shorter EPS.

To explore the relationship between note-taking and interpreting performance, a group of three raters evaluated the performances. The purpose of evaluation is to differentiate the performances as finely as possible, so that the relationship between note-taking and interpreting performance could be revealed. What is needed here is a stringent rating system to differentiate the performances of professional interpreters. Accuracy is a core component of interpreting quality (e.g., Gile, 1999; Pöchhacker, 2002) and is applied in many studies to quantify interpreting performance (e.g., Dam & Engberg, 2006; Gerver, 1969/2002; Liu & Chiu, 2009).

For studies on note-taking, accuracy is "particularly relevant and central" because the purpose of note-taking is to ensure an accurate interpretation (Dam & Engberg, 2006, p. 216). Therefore, the current study used a proposition-based rating method, using accuracy as the only criterion. With the proposition analysis results provided by *CPIDR 5*, the Chinese speech was divided into 101 scoring units, and the English speech was divided into 112 scoring units, with each scoring unit containing an average of three propositions. Raters were given the source and target texts in parallel to each other with the scoring units marked onto them. Each unit was given a score of either 1 or 0 depending on how closely the target text unit matches the source. The percentage of units correctly interpreted was used as the final score to indicate the interpreting performance. More details about how the scoring units are determined can be found in (Chen, 2017).

Each interpreting performance was rated by all three raters. A high degree of reliability was found between the raters. The average measures intraclass correlation coefficient (ICC) in the E-C task was .95 with a 95% confidence interval from .69 to .98; $F(21, 42) = 52.2$, $p < .001$. The ICC in the C-E task was .94 with a 95% confidence interval from .87 to .98; $F(21, 42) = 20.7$, $p < .001$. The mean of the three ratings was used as the score of the interpreting performance.

## 5.4 Results

### 5.4.1 What do interpreters note down

A paired-samples t-test was conducted to compare the total number of note units taken down in the two tasks. No significant difference was found between the E-C task ($M$ = 182.1, $SD$ = 27.7) and the C-E task ($M$ = 179.3, $SD$ = 27.6); $t(21) = 0.90$, $p = .380$. This shows that on average participants took similar amounts of notes in both tasks. A further step was taken to see what proportion of the source speech has been put into notes. All note units that have a one-to-one correspondence with the source speech (about 90% of the notes) were selected and mapped onto the source speech. There was no difference between the quantities of notes selected in the two tasks. On average, the interpreters put about one third of the source speech elements in notes, but there was a difference between the two tasks. The proportion of source speech noted down in the E-C task was 27.4% ($SD$ = 4.16%), significantly lower than the 33.8% that was noted down in the C-E task ($SD$ = 5.0%); $t(21) = -9.12$, $p < .001$, $d = 1.94$[8].

The frequencies of different categories of source speech units to be noted down by interpreters are shown in Table 5.2. Results show that the frequency of numbers (including dates) being noted down is the highest in both tasks (over 95%), followed by proper nouns, adjectives, and nouns (between 50% and 70%). Adverbs, verbs, and conjunctions in both tasks, along with localisers and pronouns in the C-E task, were noted down in between 15% and 40% of the frequencies. Noted down at the lowest frequencies (below 10%) were: determiners and prepositions (in both tasks), modal verbs and pronouns (in the E-C task), and measure words (in the C-E task). Additionally, each task contains two lists. The frequency of each list being noted down was 100% (although not all participants noted down the complete list).

[8] Cohen's $d$ is the effect size used in this article to compare between two means. It is calculated as the difference between the means divided by the pooled standard deviation. Effect sizes are classified as small ($d$ = 0.2), medium ($d$ = 0.5), and large ($d$ = 0.8).

Table 5.2 Categories of the source speech units and their frequencies of being noted down

| E-C task | | C-E task | |
|---|---|---|---|
| Number | 98.4% | Number | 97.9% |
| Proper noun | 68.4% | Proper noun | 69.2% |
| Adjective | 58.4% | Adjective | 48.1% |
| Noun | 50.9% | Noun | 43.6% |
| Adverb | 38.1% | Localiser | 37.1% |
| Verb | 21.6% | Verb | 25.5% |
| Conjunction | 14.7% | Conjunction | 23.3% |
| Preposition | 9.0% | Adverb | 16.8% |
| Modal | 2.4% | Pronoun | 15.9% |
| Pronoun | 1.9% | Determiner | 8.2% |
| Determiner | 0.8% | Preposition | 6.1% |
| | | Measure word | 5.2% |

### 5.4.2 When do interpreters take notes

Interpreters were engaged in simultaneous listening (and analysis) and note-writing for about the same percentages of time in the E-C task ($M$ = 56.9%, $SD$ = 6.08%) and the C-E task ($M$ = 56.4%, $SD$ = 5.27%); there was no significant difference between the tasks, $t(21) = 0.37$, $p = .715$. That is to say, regardless of the direction of interpreting, interpreters spent about 60% of the time on parallel processing of multiple tasks in the first phase of CI.

The time lag between listening and note-writing, namely the EPS, was 2447 milliseconds averaged across the two directions of interpreting (2262 milliseconds in the E-C task and 2632 milliseconds in the C-E task). The EPS of numbers, which were treated as one special category of notes (see Section 3.5.3), was of particular interest. The EPS of number notes in the E-C and C-E task was 1005 milliseconds ($SD$ = 964) and 1794 milliseconds ($SD$ = 583) respectively, both significantly shorter than the average EPS. This means that interpreters reduced the EPS when they come across numbers and started taking notes as soon as they heard one.

### 5.4.3 How do interpreters take notes

On the choice of form between language and symbol, interpreters preferred language to symbol in both tasks (Table 5.3). In the E-C task, interpreters took 75.2% notes in language, significantly higher than the 19.8% in symbol. In the C-E task, interpreters took 68.6% notes in language, also significantly higher than the 22.5% in symbol. On the choice of form between abbreviation and full word, no significant difference was found between the distributions of the two note forms in either task (Table 5.3).

Table 5.3 The choice of form

| | Language | | Symbol | | | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | 75.2% | 7.34% | 19.8% | 7.59% | 17.42*** | 21 | 3.71 |
| | Abbreviation | | Full word | | | | |
| | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| | 39.6% | 10.2% | 35.5% | 12.0% | 0.91 | 21 | |
| | Language | | Symbol | | | | |
| | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| C-E | 68.6% | 8.11% | 22.5% | 8.54% | 13.1*** | 21 | 2.79 |
| | Abbreviation | | Full word | | | | |
| | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| | 37.6% | 7.27% | 31.0% | 10.4% | 1.94 | 21 | |

*** $p < .001$

On the choice of language, interpreters preferred English to Chinese in both tasks (Table 5.4). In the E-C task, interpreters took 16.4% notes in Chinese, significantly less than the 58.8% in English. In the C-E task, interpreters took 26.7% notes in Chinese, also significantly less than the 41.9% in English. There were also differences between the two tasks. The percentage of Chinese notes in the E-C task (16.4%) was significantly lower than that in the C-E task (26.7%), $t(21) = -2.60$, $p = .017$, $d = 0.55$; the percentage of English notes in the E-C task (58.8%) was significantly higher than that in the C-E task (41.9%), $t(21) = 4.16$, $p < .001$, $d = 0.89$. That is to say, interpreters' choice of language (preferring English over Chinese) was affected by the direction of interpreting, and to be more specific, by the source/target status of the languages. When English and the source language coincided, the preference for English was strengthened, but when English and the source language contradicted (when Chinese was the source language), this preference was weakened.

Table 5.4 The choice of language

| | Chinese | | English | | | | |
|---|---|---|---|---|---|---|---|
| Task | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | 16.4% | 13.4% | 58.8% | 15.2% | -7.15*** | 21 | 1.52 |
| C-E | 26.7% | 16.4% | 41.9% | 16.5% | -2.24* | 21 | 0.48 |

* $p < .05$, *** $p < .001$

### 5.4.4 Why do interpreters make different note-taking choices

#### 5.4.4.1 The physical, temporal and cognitive demands of different note-taking choices

Significant differences were found between language and symbol notes in terms of the distance and duration of pen and the EPS in both directions of interpreting (Table 5.5). The distance and duration of language notes were significantly longer than those of symbol notes, and EPS of language notes was significantly shorter than that of symbol notes. This shows that the physical and temporal demands of language notes were significantly higher than those of symbols, but the cognitive load associated with language notes was significantly lower than symbols. The fact that interpreters preferred language over symbol and that the cognitive load associated with language notes was lower than symbol notes (even though the physical and temporal demands of the former were higher) seemed to indicate that the choice between language and symbol was governed by the cognitive demand (rather than the physical or temporal demands).

Table 5.5. A comparison of language and symbol notes in terms of distance, duration and EPS

| | | Language | | Symbol | | | | |
|---|---|---|---|---|---|---|---|---|
| Task | Pen data | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | Distance (cm) | 7.87 | 2.13 | 3.18 | 1.12 | 13.3*** | 21 | 2.84 |
| | Duration (ms) | 1214 | 209 | 378 | 63 | 20.0*** | 21 | 4.26 |
| | EPS (ms) | 2282 | 610 | 2694 | 991 | -3.41** | 21 | 0.73 |
| C-E | Distance (cm) | 8.14 | 2.28 | 3.10 | 1.14 | 12.3*** | 21 | 2.61 |
| | Duration (ms) | 1236 | 218 | 365 | 91 | 17.4*** | 21 | 3.70 |
| | EPS (ms) | 2653 | 695 | 3047 | 741 | -4.33*** | 21 | 0.92 |

** $p < .01$, *** $p < .001$

Significant differences were found between abbreviation and full word notes in terms of the distance and duration of pen in both directions of interpreting, but not in the EPS (Table 5.6). The distance and duration of abbreviation notes were significantly shorter than those of full word notes, but the EPS was similar between the two. That is to say, the physical and temporal demands of abbreviations were significantly lower than those of full words, but the cognitive load was similar. The fact that interpreters did not have a preference between abbreviation and full word and that the levels of cognitive load associated with the two forms of notes were similar (even though the physical and temporal demands of the former were lower) seemed to indicate that the choice between abbreviation and full word was also governed by the cognitive demand (rather than the physical or temporal demands).

Table 5.6 A comparison of abbreviation and full word notes in terms of distance, duration and EPS

| | | Abbreviation | | Full word | | | | |
|---|---|---|---|---|---|---|---|---|
| Task | Pen data | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | Distance (cm) | 6.74 | 1.92 | 9.17 | 2.54 | -7.27*** | 21 | 1.55 |
| | Duration (ms) | 1068 | 212 | 1384 | 232 | -6.61*** | 21 | 1.41 |
| | EPS (ms) | 2338 | 633 | 2203 | 629 | 1.62 | 21 | |
| C-E | Distance (cm) | 7.12 | 1.93 | 9.40 | 2.86 | -6.57*** | 21 | 1.40 |
| | Duration (ms) | 1080 | 178 | 1423 | 289 | -6.78*** | 21 | 1.45 |
| | EPS (ms) | 2660 | 737 | 2627 | 700 | 0.43 | 21 | |

*** $p < .001$

No significant difference was found between Chinese and English notes in terms of the distance and duration of pen in either direction of interpreting, but the EPS of Chinese notes was significantly longer than that of English notes in the C-E task (but not in the E-C task) (Table 5.7). This means that the physical and temporal demands of different language choices were similar. However, the cognitive load associated with Chinese notes was significantly higher than that with English notes in the C-E task. It would seem that in L1-L2 interpreting, the choice of language was determined by the cognitive demand (rather than the physical or temporal demands) associated with different language choices. But in L2-L1 interpreting, the choice of language was determined by the source/target language status and was dominated by the source language.

Table 5.7 A comparison of Chinese and English notes in terms of distance, duration and EPS

| Task | Pen data | Chinese *M* | Chinese *SD* | English *M* | English *SD* | *t* | *df* | *d* |
|---|---|---|---|---|---|---|---|---|
| E-C | Distance (cm) | 7.61 | 1.91 | 8.04 | 2.22 | -1.24 | 21 | |
| | Duration (ms) | 1199 | 444 | 1189 | 187 | 0.13 | 21 | |
| | EPS (ms) | 2368 | 766 | 2213 | 576 | 1.68 | 21 | |
| C-E | Distance (cm) | 8.10 | 2.81 | 8.22 | 2.17 | -0.35 | 21 | |
| | Duration (ms) | 1204 | 245 | 1242 | 218 | -0.75 | 21 | |
| | EPS (ms) | 2859 | 857 | 2611 | 669 | 2.71* | 21 | 0.58 |

* $p < .05$

#### 5.4.4.2 Note-taking and interpreting performance

The relationship between interpreting performance and note-taking was investigated using the Pearson's correlation (Table 5.8). The performance was positively correlated with the total number of note units taken in the C-E task but not in the E-C task. There was a positive correlation between the interpreting performance and the percentage time spent on note-taking (i.e., engaged in simultaneous listening/analysis and note-writing) in both tasks. The performance was negatively correlated with the EPS in the E-C task but not in the C-E task.

The performance was also correlated with the distribution of notes, i.e., the percentage of notes in different categories. In the E-C task, there was a negative correlation between the performance and the percentage of language notes and a positive correlation between the performance and the percentage of symbol notes, but no significant results was found in the C-E task. There was a negative correlation between the interpreting performance and the percentage of English notes in both the E-C task and the C-E task. No significant correlation was found between the performance and the percentage of Chinese, abbreviation, and full word notes.

Table 5.8 The Pearson's correlation between interpreting performance and note-taking

| Task | Quantity of notes | Time spent on note-taking | EPS | Percentage of notes: Language | Percentage of notes: Symbol | Percentage of notes: English |
|---|---|---|---|---|---|---|
| E-C | | .52* | -.44* | -.55** | .56** | -.48* |
| C-E | .48* | .50* | | | | -.43* |

$N = 22$, * $p < .05$, ** $p < .01$

As we can see, the two directions of interpreting shared some common trends. A higher percentage of time spent on multi-tasking (listening/analysing and note-taking at the same time) was related to a better interpreting performance, but a higher percentage of English notes was related to a worse interpreting performance. However, there were also some differences between the two directions. In L2-L1 interpreting, the interpreting performance was better when the EPS was shorter, when interpreters used more symbol notes and when they used less language notes. But in L1-L2 interpreting, the quality of performance was better when interpreters took more notes.

## 5.5 Discussions

### 5.5.1 The content of note-taking

The empirical data in this study supports most (but not all) of what has been suggested in literature on what should be noted down (see Section 2). According to the results, interpreters give top priority to lists (100% noted down), numbers and dates (almost 100% noted down), and proper names (about 70% noted down). In the "Subject Verb Object" unit, the frequency of nouns (usually subjects or objects) being noted down is about 50%; the frequency of verbs being noted down is about 20%. The frequency of conjunctions (usually links) being noted down is also about 20%.

Standing in contrast to what has been recommended in literature, verb tenses and modal verbs are rarely noted down (less than 5%). This is likely to due to the features of the specific language pair of Chinese and English. Marking verb tenses is not always necessary (especially in the E-C direction) because the form of a Chinese verb never changes; rather, tenses in Chinese are inferred from context or marked by temporal words (Liu, 2008, p. 69). As to modal verbs, the Chinese modals are functionally weak compared with English, and Chinese does not differentiate all the meanings associated by the English modals (Li, 2010, p. 359). These differences between Chinese and English could probably explain why the data in this study partly contradicts the prescriptive studies, many of which do not involve Chinese.

When it comes to interpreter education, students should be trained to pay special attention to lists, numbers and proper names, and try to note down these elements as completely as possible. Students should also be trained to analyse the "Subject Verb Object" structures in the source speech, locate the main ideas and note them down. Teachers should also be careful with the unique features brought by specific language pairs and make pedagogical adaptations accordingly, rather than strictly follow the established note-taking systems.

### 5.5.2 The timing of note-taking

On average, interpreters lag behind the source speech for about 2.45 seconds during the listening and note-taking phase of CI. This is shorter than what Andres (2002) found in her study, which was between 3 and 6 seconds. The difference could be caused by the method of calculating the span. Andres calculated the span from the start of sound and manually determined the start of note-taking by checking video recordings. This study calculated the EPS from the end of sound, and the start of note-taking was automatically reported by software. Another possible cause for the difference is the language pair. In studies on simultaneous interpreting, the reported EVS in interpreting between different languages pairs varies (e.g., Barik, 1973; Lee, 2002; Oléron & Nanpon, 1965/2002; Timarová et al., 2011; Treisman, 1965).

One particularly interesting finding is that interpreters note down numbers exceptionally quickly. The EPS data shows that numbers are noted down about 1 second faster than language notes and about 1.5 seconds faster than symbols. It would seem that in order to avoid cognitive overload (numbers are a well-known problem-trigger in interpreting (Alessandrini, 1990; Cheung, 2008)), interpreters opted for a strategy to shorten the time lag and lower the cognitive load when they come across numbers.

Since the data in this study is gathered from professional interpreters, it could serve as a reference as to how much interpreters should lag behind the source speech in note-taking. The time lag found in students' interpretations could be checked against the data to get an idea of whether students are following the source speech too closely or lagging too far behind. In particular, students should be instructed to keep the time lag short and take down notes as quickly as possible when it comes to numbers.

### 5.5.3 The choice of form and the choice of language

Interpreters showed a clear preference for language over symbol, corroborating the previous studies (Andres, 2002; Dai & Xu, 2007; Dam, 2004a, 2004b; Lung, 2003). No preference was found between abbreviation and full word. Previous studies did not perform significance testing on their data, but the descriptive data showed a preference for abbreviations in some studies (e.g., Dai & Xu, 2007) but full words in others (e.g., Dam, 2004a; Lung, 2003). The inconsistent findings could be related to such factors as the language pair and the type of participants used (professional vs. student interpreters), but there is not enough empirical evidence to draw conclusions.

Interpreters preferred English (L2) in both directions of interpreting, and this preference was stronger when English was the source language. The result contradicts what Dam (2004b) found in students with the Danish/Spanish language pair, in which all participants preferred their L1, the better mastered native language. The result is in line with what Szabó (2006) found in professional interpreters with the Hungarian/English language pair. Szabó observed a preference for English (L2), regardless of the direction of interpreting, and pointed to the morphological complexity and economy of writing as an explanation.

A post hoc questionnaire of this study shows that interpreters prefer to use English in note-taking mainly because English can be written down using phonetic spelling and even misspelling, whereas the written form of Chinese is not always available immediately. The participants find it easier to take notes in English. This has to do with the fact that English is an alphabetic language but Chinese is not. However, it should also be noted that the participants in this study are based in Australia, an English-speaking country, and they are likely to have a very strong L2, same as the participants in Szabó's study.

Findings in this study and various previous studies show that the choice of language in note-taking is a function of the combined influence from a series of factors, including: (1) the language combination itself (e.g., alphabetic/non-alphabetic nature; morphological complexity); (2) the L1/L2 language status; (3) the source/target language status; and (4) interpreter characteristics (e.g., working experience and language competence).

### 5.5.4 Note-taking, cognitive load and interpreting performance

The time needed for writing, the effort of the hand and the mental effort costs are believed to be the main concerns in making note-taking choices (e.g., Alexieva, 1994, pp. 203-204; Allioni, 1989, p. 195; Gile, 2009, p. 178). But the data in this study suggests that the temporal and physical demands are less of a concern to interpreters than the cognitive demand. This is evident in the different note-taking choices made by interpreters (see Table 5.9). Although the physical and temporal demands of language notes are higher than those of symbols, the cognitive demand of language notes is lower and interpreters preferred language to symbol. The cognitive demands of abbreviation and full word notes are similar and interpreters do not have a preference between the two, even though abbreviations are easier and faster to write. Chinese and English notes are similar in their physical and temporal demands, but the cognitive load associated with English is lower and interpreters prefer English in the C-E task[9] (although it is the non-source and non-native language).

However, the relation between note-taking and interpreting performance seems to show that the choices made by interpreters to lower cognitive load in the listening and note-taking phase of CI are at the expense of interpreting quality (see Table 5.9). A better interpreting performance is correlated with a lower percentage of language notes and a higher percentage of symbol notes (only the E-C task reached significance); it is also correlated with a lower percentage of English notes (but not with Chinese notes).

Table 5.9 The physical, temporal and cognitive demands and interpreters' preferred note-taking choices

| | Choice of form | | Choice of language |
|---|---|---|---|
| | Language vs. symbol | Abbreviation vs. full word | Chinese vs. English |
| Physical and temporal demands | Language > Symbol | Abbreviation < Full word | Chinese ≈ English |
| Cognitive load | Language < Symbol | Abbreviation ≈ Full word | Chinese > English (C-E only) |
| Preferred note-taking choice | Language | No preference | English |
| Relation with performance | Less language, more symbol, better performance (E-C only) | No relation | Less English, better performance; no relation with Chinese |

Considering the positive contribution of symbols to interpreting performance, more emphasis should be put on symbols in CI training. The aims should be two-fold: to decrease the cognitive costs of symbols and to increase their usage. Instead of merely pointing to the highly individualised nature of symbols in note-taking and leaving it to the students to develop a system of their own (usually painstaking and involving a lot of trial and error), teachers could lend a set of symbols rich enough for the students to start with. The students could then spend more time on practising using the symbols instead of inventing them.

[9] In the E-C task there is no difference in cognitive load and the language choice seems to be determined by the source language status of English.

Some important differences exist between the two directions of interpreting. In the L2-L1 direction, the performance is negatively correlated with the EPS, whereas in the L1-L2 direction, the performance is positively correlated with the quantity of the notes. This suggests that when interpreting from the non-native language, the time lag matters more. Because the cognitive resources needed for listening and analysis in the non-native language is higher, interpreters need to maintain a shorter EPS to ensure that they do not overload their cognitive system. But when interpreters work from the native language, the cognitive demands of listening and analysis is lower. Therefore, they can afford to allocate more resources to note-taking and take down more notes, which in turn stores more information and facilitates a complete production.

The findings on directionality could inform teachers to take different pedagogical approaches to the two directions of interpreting and remind students to develop different strategies to deal with the specific challenges.

## 5.6 Conclusion

This study contributes a process-oriented approach towards note-taking and CI. It makes a methodological contribution by showing how digital pen recording can be applied in interpreting research. Pen recording is found to be a powerful tool to tap into the cognitive processes in note-taking and CI. The data not only provides us with an accurate and encompassing picture of the interpreting process but also serves as an indicator of cognitive load.

The study conducts a comprehensive investigation on the note-taking choices, cognitive load and interpreting performance in CI. Some fundamental principles underlying the interpreters' note-taking choices (which seem highly individualised at a first glance) are unveiled and pedagogical suggestions are made based on empirical data.

It has to be admitted that this study is limited in several ways. Only one language pair (Chinese and English) is involved, so the findings may not generalise to another language combination. Most of the participants live and work in a country where English (their L2) is spoken, and the results may not generalise to another type of interpreters, such as Chinese-English interpreters based in China. Only one group of participants is involved in the study. This design does not allow the investigation of the role interpreting experience plays in the process. It would be interesting to see how different groups of participants with varying degrees of experience differ in the note-taking and CI process. These limitations have pointed to some interesting directions for future research. It is hoped that more researchers will be joining the effort to create a more comprehensive picture of note-taking and CI.

## 5.7 References


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting, 14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Albl-Mikasa, M. (2006). Reduction and expansion in notation texts. In C. Heine, K. Schubert, & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp. 195–214). Tübingen: Narr.

Albl-Mikasa, M. (2008). (Non-) sense in note-taking for consecutive interpreting. *Interpreting, 10*(2), 197–231. doi:10.1075/intp.10.2.03alb

Alessandrini, M. S. (1990). Translating numbers in consecutive interpretation: An experimental study. *The Interpreters' Newsletter, 3*, 77–80.

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199–206). Amsterdam: John Benjamins.

Allioni, S. (1989). Towards a grammar of consecutive interpretation. In L. Gran & J. M. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation* (pp. 191–197). Udine: Campanotto.

Andres, D. (2002). *Konsekutivdolmetschen und notation [Consecutive interpreting and note-taking]*. Frankfurt: Peter Lang.

Barik, H. C. (1973). Simultaneous interpretation: Temporal and quantitative data. *Language and Speech, 16*(3), 237–270. doi:10.1177/002383097301600307

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40*(2), 540–545. doi:10.3758/BRM.40.2.540

Cardoen, H. (2013). The Effect of Note-taking on Target-text Fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging Research in Translation Studies: Selected Papers of the CETRA Research Summer School 2012* (pp. 1-22). Leuven: CETRA.

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

Chen, S. (2017). Note-taking in consecutive interpreting: New data from pen recording. *Translation and Interpreting, 9*(1), 4–23. doi:10.12807/ti.109201.2017.a02

Cheung, A. K. F. (2008). Simultaneous interpreting of numbers: An experimental study. *Forum, 6*(2), 23–38. doi:10.1075/forum.6.2.02kfc

Christoffels, I. K., & De Groot, A. M. B. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition, 7*(3), 227–240. doi:10.1017/S1366728904001609

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research, 39*(2), 136–144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting, 6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters’ notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V., & Engberg, J. (2006). Assessing accuracy in consecutive interpreting: a comparison of semantic network analyses and intuitive assessments. In C. Heine, K. Schubert, & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp. 215–234). Tübingen: Narr Francke Attempto Verlag.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (Vol. 7, pp. 227–254). Berlin: Mouton de Gruyter.

Ehrensberger-Dow, M., Dimitrova, B. E., Hubscher-Davidson, S., & Norberg, U. (Eds.). (2015). *Describing cognitive processes in translation: Acts and events*. Amsterdam: John Benjamins.

Gerver, D. (1969/2002). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 53–66). London and New York: Routledge.

Gile, D. (1991). Prise de notes et attention en début d'apprentissage de l'interprétation consécutive–une expérience– démonstration de sensibilisation [Note-taking and attention at the beginning of consecutive interpreting learning –an experience– demonstration of awareness]. *Meta, 36*(2–3), 431–439.

Gile, D. (1999). Variability in the perception of fidelity in simultaneous interpretation. *Hermes, 22*, 51–79. doi:10.7146/hjlcb.v12i22.25493

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

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. Manchester and Kinderhook: St. Jerome.

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

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation, 6*, 53-77.

Ilg, G. (1988). La prise de notes en interprétation consécutive. Une orientation générale [Note-taking in consecutive interpretation: A general approach]. *Parallèles, 9*, 9–13.

Jones, R. (1998). *Conference interpreting explained*. Manchester: St. Jerome.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Lee, T.-H. (2002). Ear voice span in English into Korean simultaneous interpretation. *Meta, 47*(4), 596-606. doi:10.7202/008039ar

Li, R. (2010). A contrastive study of the semantic functions and pragmatic uses of modal verbs in English and Chinese. In D. Shu & K. Turner (Eds.), *Contrasting meaning in languages of the East and West* (pp. 335-360). Bern: Peter Lang.

Liu, M. (2008). *逐步口譯與筆記 [Consecutive interpretation and note-taking]*. Taipei: Bookman.

Liu, M., & Chiu, Y.-H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting, 11*(2), 244–266. doi:10.1075/intp.11.2.07liu

Lung, R. (1999). Note-taking skills and comprehension in consecutive interpretation. *Babel, 45*(4), 311–317. doi:10.1075/babel.45.4.04lun

Lung, R. (2003). Taking "notes" seriously in the interpretation classroom. In Á. Collados Aís, M. M. Fernández Sánchez, & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación [The evaluation of quality in interpretation: Research]* (pp. 199–205). Granada: Comares.

Martín, R. M. (Ed.) (2016). *Reembedding translation process research*. Amsterdam: John Benjamins.

Matyssek, H. (1989). *Handbuch der notizentechnik für dolmetscher [Handbook of note-taking for interpreters]*. Heidelberg: Julius Groos.

Oléron, P., & Nanpon, H. (1965/2002). Research into simultaneous translation. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 43–50). London and New York: Routledge.

Orlando, M. (2010). Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter, 15*, 71–86.

Pöchhacker, F. (2002). Researching interpreting quality: Models and methods. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21st century: Challenges and opportunities* (pp. 95–106). Amsterdam: John Benjamins.

Rozan, J.-F. o. (2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Cracow: Tertium Society for the Promotion of Language Studies.

Santorini, B. (1990). *Part-of-speech tagging guidelines for the Penn Treebank Project (3rd revision)*. Retrieved from University of Pennsylvania: http://repository.upenn.edu/cis_reports/570

Schweda-Nicholson, N. (1993). An introduction to basic note-taking skills for consecutive interpretation. In E. Losa (Ed.), *Keystones of communication: Proceedings of the 34th Annual Conference of the American Translators Association* (pp. 197–204). Medford: Learned Information.

Seleskovitch, D. (1975). *Langage, langues et mémoire: étude de la prise de notes en interprétation consécutive [Speech, language, and memory: A study of note-taking in consecutive interpreting]*. Paris: Minard Lettres Modernes.

Seleskovitch, D., & Lederer, M. (1989/1995). *A systematic approach to teaching interpretation* (J. Harmer, Trans.). Silver Spring, MD: Registry of Interpreters for the Deaf.

Shreve, G. M., & Angelone, E. (Eds.). (2010). *Translation and cognition*. Amsterdam and Philadelphia: John Benjamins.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting, 8*(2), 129–147. doi:10.1075/intp.8.2.02sza

Timarová, Š., Dragsted, B., & Hansen, I. G. (2011). Time lag in translation and interpreting: A methodological exploration. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 121–146). Amsterdam and Philadelphia: John Benjamins.

Tirkkonen-Condit, S., & Jääskeläinen, R. (Eds.). (2000). *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research*. Amsterdam and Philadelphia: John Benjamins.

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology, 56*(4), 369–379. doi:10.1111/j.2044-8295.1965.tb00979.x

Xia, F. (2000). *The part-of-speech tagging guidelines for the Penn Chinese Treebank (3.0)* (38). Retrieved from University of Pennsylvania: http://repository.upenn.edu/ircs_reports/38

## An introductory note to Chapter 6

Following the analysis of pen and voice data in Chapter 5, Chapter 6 reports the eye tracking data collected in the main study of the PhD project. The focus is on Phase II of CI in which interpreters read back their notes and produce a target speech. The chapter contributes a discussion on the eye movement measures used in the study, accompanied by a definition to each measure and examples of application. It demonstrates how eye tracking can be applied in interpreting research, especially research involving note-taking and CI.

Through a combined analysis of the qualitative and quantitative data, Chapter 6 presents how visual processing is carried out during note-reading. It investigates the cognitive load in the interpreting process using a series of eye movement measures. In addition, Chapter 6 puts two pieces of the puzzle together by synthesising the findings on both Phase I and Phase II of CI. It reveals that there is a trade-off between the cognitive costs of different note-taking choices in the two phases, and further discloses the interplay between note-taking, cognitive load and interpreting performance.

# Chapter 6 An eye-tracking approach to note-reading in consecutive interpreting: Reading patterns and cognitive load[1]

**Abstract：** This study reports the eye tracking data collected from professional interpreters while they perform consecutive interpreting (CI) with notes. It is among the first to visualise the way in which note-reading proceeds. Data collected in this study provide important indicators of cognitive processing and cognitive load in the interpreting process. It then transpires that the note-taking choices made during Phase I of CI (in which interpreters listen to the source speech and write notes) affect the level of cognitive load in Phase II (in which interpreters read back their notes and produce a target speech). The results indicate that there is a trade-off between the cognitive costs in Phase I and Phase II, with interpreters strategically lowering the load in Phase I (during which interpreters are dealing with the input information for the first time and are paced by the speaker) and leaving the burden of increased cognitive load to Phase II (during which interpreters are more familiar with the information content and set their own pace).
**Keywords:** eye tracking; note-reading; note-taking; consecutive interpreting; cognitive load

## 6.1 Introduction

Translation and interpreting process research has made important progress in the past two decades. However, in terms of interpreting, simultaneous interpreting (SI) has received the bulk of the attention, with little attention devoted to studying the process of consecutive interpreting (CI). It may be the case that CI has been largely replaced by SI in some markets (such as western Europe) and contexts (such as multilateral and multilingual conferences), but it remains an important mode of interpreting in other markets (such as Asia) and the preferred mode in various other contexts such as diplomatic and business negotiations (Andres, 2015; Dam, 2010). CI is also an important part in the curriculum of interpreter training (Setton & Dawrant, 2016, pp. 82-83). Therefore, it is somewhat surprising that CI has been left behind in the vigorous development of process-oriented research on translation and interpreting.

[1] This chapter has been prepared for publication as: Chen, S. (manuscript prepared for publication). An eye-tracking approach to note-reading in consecutive interpreting: Reading patterns and cognitive load.

CI is an interesting activity from both a cognitive and a linguistic point of view. Similar to SI, it requires a high level of bilingual language processing and challenges the interpreter's cognitive system by requiring multi-tasking under strict time constraints. But CI[2] also introduces a new challenge: note-taking. In addition to listening to the source speech and producing a target speech, CI requires the interpreter to perform the tasks of note-writing and note-reading and to deliver the output relying on memory with the assistance of notes. In Phase I of CI, interpreters listen to and analyse the source speech, keep parts of the speech in their working memory, and write down notes. In Phase II, interpreters read back their notes, retrieve information from their working memory, and produce a target speech.

Phase I of CI has received some attention, with most of the efforts devoted to note-taking and in particular its product, the notes (see Chen (2016) for a review). Phase II of CI, however, has been under-researched relative to its importance to professional practice and its potential for providing insights into cognitive processes. It is evident that researchers either stop at observing the products of note-writing or shift their attention from written notes directly to interpreting performance, ignoring the actual processes involved in the note-reading and production phase. This gap in the literature may very well be due to the fact that process-oriented research on CI has been impeded by inadequate research methods.

Digital pen recording is a valuable research method for Phase I, but it has not been applied in interpreting research until quite recently (Chen, 2017b; Orlando, 2010, 2014). Before that, process research on CI relied on less accurate and more cumbersome methods such as video recording (which involved manually checking the timing of note-writing) and even then, very little can be found in the literature (see however Andres (2002)). Another research method with substantial potential for investigating the cognitive processes in Phase II of CI is eye tracking (Chen, 2017a), a method that has long been applied in fields such as reading, scene perception, and cognitive sciences, but only became a popular research method in translation studies during roughly the past decade (Hvelplund, 2017).

This study uses eye tracking to study the process of CI. It aims to make a methodological contribution to interpreting research, especially process-oriented research. The focus of the study is Phase II of CI, in which interpreters read back their

[2] In this article, CI refers to long consecutive in which systematic note-taking is used.

notes and produce a translated speech. Through an analysis of the eye movement data, the study is among the first to visualise the way note-reading proceeds in CI and provides us some insight into interpreters' "black box" during this complex cognitive and linguistic process.

### 6.2 Note-taking and note-reading in CI

Note-taking in CI is a particularly interesting topic in interpreting research. Both prescriptive and descriptive accounts of note-taking have been generated over the years. On the prescriptive side, various note-taking systems have been established in a substantial volume of literature, including those that address note-taking as a subject in its own right (e.g., Gillies, 2005; Kirchhoff, 1979; Matyssek, 1989; Rozan, 1956) and those that include note-taking as part of their discussions on interpreter training and education (e.g., Bowen & Bowen, 1980; Ilg & Lambert, 1996; Jones, 1998; Seleskovitch & Lederer, 1995). A series of skill-oriented issues have been discussed, such as how to arrange the layout of the notes, how to use symbols and abbreviations, and which language to use (source or target language) for taking notes.

On the descriptive side, note-taking in CI has been the focus of a number of empirical studies. Important variables investigated include the choice of form (e.g., Dam, 2004a), the choice of language (e.g., Abuín González, 2012; Dam, 2004b; Szabó, 2006), and the relationship between note-taking and interpreting performance (e.g., Cardoen, 2013; Dai & Xu, 2007; Dam, 2007). In terms of the choice of form, interpreters can take notes either in language or in symbols; if they take notes in the form of language, they could use either abbreviations or full words. In terms of the choice of language, interpreters can choose between the source and the target language, or between the native and non-native language.

Findings in those empirical studies have pointed to some general trends. Interpreters take much more notes in language than in symbol, but no consistent preference for abbreviation or full word has been found. Their choice of language in note-taking is dependent on a series of factors, such as the nature of the languages themselves (e.g., morphological complexity and economy of writing), the native/non-native language status, the source/target language status, task characteristics (e.g., texture or genre of the source speech) and interpreter characteristics (e.g., working experience and language competence).

Compared to the amount of research efforts devoted to note-taking in Phase I of CI, the attention granted to note-reading in Phase II has been very limited. Note-reading is in fact a very important process in CI. After interpreters have listened to the source speech and taken down notes, they read back their notes and produce a translated speech. The reading process takes place simultaneously with the speaking process and provides information for the interpretation to cue and supplement the working memory. Poor quality notes have a detrimental effect on production because too much effort will be allocated to note-reading. Good quality notes, on the other hand, can facilitate target speech production, "telling the interpreter when to pause, when to add emphasis and when not to" (Gillies, 2005, p. 7).

Moreover, the note-taking choices made by interpreters during Phase I can have a direct bearing on Phase II of CI. A recent study (Chen, 2017c) has shown that the varied note-taking choices made in Phase I are associated with different levels of cognitive load, and that interpreters' note-taking choices are mainly determined by cognitive rather than physical and temporal demands (Table 6.1). However, very little is known about how these varied note-taking choices made in Phase I affect the cognitive processing and cognitive load in Phase II. This is one of the questions that the current article will attempt to answer.

Table 6.1 Note-taking choices and cognitive load in Phase I of CI

| | Choice of form | | Choice of language |
|---|---|---|---|
| | Language vs. symbol | Abbreviation vs. full word | Chinese vs. English |
| Physical and temporal demands | Language > Symbol | Abbreviation < Full word | Chinese ≈ English |
| Cognitive load in Phase I | Language < Symbol | Abbreviation ≈ Full word | Chinese > English (C-E only) |
| Preferred note-taking choice | Language | No preference | English |

Adapted from (Chen, 2017c)

What makes note-reading even more intriguing is that it is a very unique reading process. Interpreters do not "read" their notes in the usual sense of the word, so note-reading both resembles and differs from reading in general. Similar to general reading, interpreters need to read the notes to grab information. Therefore, in terms of information intake, note-reading should not deviate too much from normal reading. But what sets note-reading apart from general reading is that interpreters are simultaneously performing the cognitive process of producing a translation of the source speech from memory supported by what they are reading. The prescriptive literature suggests that interpreters should be reading ahead and preparing the next section of the target speech while producing a translated speech for a former section (Gillies, 2005, p. 73). But so far there is little empirical evidence of this particular reading behaviour.

Eye tracking is a particularly useful tool for studying note-reading in CI. By recording the eye movements of interpreters while they are reading their notes and producing a target speech, Phase II of CI can be approached in unprecedented ways. In addition to providing information on attention allocation, eye tracking data contribute important indicators of cognitive processing, allowing us to see how the note-taking choices made by interpreters during Phase I affect the level of cognitive load in Phase II. Eye tracking also allows us to visualise the way in which note-reading progresses and find traces of how note-reading interacts with memory and target speech production.

### 6.3 Eye tracking and cognitive processing in language tasks

In the late 19th century, Javal first observed that eye movements in reading consist of a series of rapid movements (saccades) and intermittent short stops (fixations) (Huey, 1908). Since those very early days, many findings about eye movements in language processing have been reported (see Liversedge, Gilchrist, and Everling (2011) for a review). One important discovery was that information intake could only occur during fixations but not during saccades. Just and Carpenter (1980) thus formulated the influential eye–mind hypothesis. It states that there should be no appreciable lag between what is fixated and what is being processed. When people look at a word or an object, they also simultaneously process it for the duration of the fixation. In this way, eye movements open a window of opportunity for studying the cognitive processes of perception and comprehension.

Eye tracking is a technique that captures where, how and when people look using devices called eye trackers. Most of the eye trackers in use today are reliant on video-based pupil and corneal reflection (Duchowski, 2007, p. 51). The eye trackers photograph the pupil and the reflection of an infrared light source from the fovea and combine the two to determine the point of gaze (i.e., where people look). The eye image (with the pupil and corneal reflection) forms one coordinate system; the stimulus (such as a screen) forms the other coordinate system. A calibration procedure is carried out to relate one system to the other by asking the participant to fixate on specific points on the stimulus and then mapping the gaze point to that.

Once the raw eye tracking data (time-stamped points of gaze) are obtained, they can be processed to produce various events (such as fixations and saccades). There are two main types of algorithms used for event detection. Dispersion-based algorithms identify fixations by finding data samples that are close enough to one another for a predetermined minimal duration of time. Velocity-based algorithms find saccades as periods longer than a minimal period of time during which the eye velocity is below a predefined threshold. Most commercial eye trackers come with built-in software for extracting events.

After processing raw data samples into events such as fixations and saccades, a further step can be taken to define areas of interest (AOIs). AOIs are certain regions in the stimulus within which the eye movement data are of central concern to a study. Usually, research questions and hypotheses should decide what AOIs to create for the stimulus. For example, in reading research, an AOI can be a single word or a string of sentences; in scene perception research, an AOI can be a human face or a moving object. When using eye tracking to study note-reading, the unit of analysis is a note unit (or a group of notes that share one common feature). In this case, an AOI needs to be drawn for every note unit.

When the unit of analysis is as small as a word or a note unit, the appropriate measure to use is controversial (Inhoff & Radach, 1998). In this context, any single eye movement measure is a pale reflection of the reality of cognitive processing. A recommended remedy is to examine multiple measures so that inferences drawn about the cognitive processing can be as valid as possible (Rayner, 1998; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Schmauder, 1992).

In this study, a series of eye movement measures is investigated for the purpose of capturing and approximating the cognitive reality of note-reading in CI (also see Section 4.5.2). Each of these measures is defined below and a brief review is given of how they have been applied in relevant research fields.

*Dwell* refers to a visit in an AOI from entry to exit. A dwell can be measured by the dwell time, which can be based on either the raw data (the sum of all data points including fixations and saccades falling within an AOI) or fixations (the sum of all fixations in an AOI). When the unit of analysis is a word (similar to the note unit in this study), the effect of adding saccade duration is minimal because intra-word saccade durations are quite brief (Rayner, 1998). Therefore, the dwell time in this study is defined as the sum of fixations only, excluding all other events. Dwell time is a commonly applied eye movement measure in various fields, but it has been suggested by a number of researchers that the measure needs to be further refined in order to draw reasonable inferences about cognitive processing (e.g., Holmqvist et al., 2011; Inhoff & Radach, 1998). To provide a fine-tuned account of the dwell time, this study distinguishes *first-pass dwell time, second-pass dwell time* and *total dwell time*.

*First-pass dwell time* (usually termed gaze duration in reading research) refers only to the first dwell in an AOI from entry to exit. Rayner (1998) reviewed reading research and concluded that first-pass dwell time (gaze duration) was a good index both of word frequency and of comprehension processes integrating several words. The measure is proposed as a candidate for measuring early processing and object recognition (Liversedge, Paterson, & Pickering, 1998).

A measure which can be confused with the first-pass dwell time is the *first fixation duration*. This is the duration of the first fixation in an AOI. It is indicative of fast processes such as recognition and identification (Holmqvist et al., 2011, p. 385) and sensitive to cognitive difficulty experienced immediately on processing an AOI (Liversedge et al., 1998).

*Second-pass dwell time* includes all subsequent dwells (excluding the first dwell) on the same AOI. This measure is a useful indicator for global text processing (Hyönä, Lorch, & Rinck, 2003) and reflects more delayed effects in sentence processing (Murray, 2000).

*Total dwell time* subsumes the fixation time in an AOI during an entire trial. It is sensitive to slow and long-term cognitive processes (Holmqvist et al., 2011). In reading research, total dwell time was found to be sensitive to linguistic processes that operated after a word had been identified (Daneman, Reingold, & Davidson, 1995). In translation studies, total dwell time is often used as an indicator of cognitive load. Many studies found that the total dwell time on the target text was longer than that on the source text, indicating more processing efforts spent on target text production and monitoring (Carl & Kay, 2011; Dragsted, 2010; Jensen, Sjørup, & Balling, 2009). Studies also found that student translators behaved differently and spent more time looking at the source text, indicating a more effortful source text analysis (Jakobsen & Jensen, 2008; Sharmin, Špakov, Räihä, & Jakobsen, 2008). Sjørup (2008) reported that the total dwell time was longer when translators came across metaphors, suggesting that dealing with metaphors required increased cognitive processing. In audio-visual translation research, it was reported that the total dwell time on subtitles was longer for deaf and hard-of-hearing viewers compared to hearing viewers, indicating a larger effort needed to process subtitled content and more difficulty in extracting information (Szarkowska, Krejtz, Pilipczuk, Dutka, & Kruger, 2016).

Another widely used eye movement measure is the *average fixation duration*. This is the sum of the duration of all fixations divided by the number of fixations in an AOI during an entire trial. As has been demonstrated in various fields (e.g., reading, scene perception and usability research), a longer average fixation duration is usually associated with a deeper and more effortful cognitive processing (Holmqvist et al., 2011, p. 381). This measure typically varies across different tasks and stimuli. The average fixation duration is about 225 milliseconds in silent reading (for comprehension) and 275 milliseconds in oral reading (Rayner, 1998); it ranges from 205 milliseconds (Jakobsen & Jensen, 2008) to 245 milliseconds (Dragsted, 2010) in reading in preparation for translation, and from 235 milliseconds (Jakobsen & Jensen, 2008) to 252 milliseconds in sight translation/interpretation (Dragsted & Hansen, 2009).

Apart from the time-based measures, a series of count-based eye movement measures are also related to cognitive processing in language tasks. The *number of fixations* refers to the fixation count inside an AOI during a trial. In translation research, the number of fixations has been found to index cognitive load. Doherty, O'Brien, and Carl (2010) provided participants with machine translated sentences with good and bad acceptability. They found that there were more fixations and longer dwell times on “bad” sentences than on “good” sentences, indicating that the former was more difficult to process. Dragsted (2012) provided words with high versus low target text variability (number of alternative renditions in the target text) to participants and found that the number of fixations on high variation words was higher than that on low variation words.

*Revisit* is defined as a transition to an AOI already visited. In picture viewing, the *number of revisits* is an indicator of a semantically informative AOI (Loftus & Mackworth, 1978) or a need to confirm (Mello-Thoms et al., 2005). In reading research, the number of revisits is related to incomplete lexical processes (Pollatsek & Rayner, 1990; Pynte, 1996).

*Regressions* are right-to-left movements along the line of reading or movements back to previously read lines (Rayner, 1998). It has been found that people make more regressions when the reading task was cognitively demanding (Frazier & Rayner, 1982). The *regression rate* reported in reading research (reading in native language) is about 10–15% (Rayner, 1998); in translation studies, the regression rate is 20–25% in reading a foreign language and 30–35% in sight translation/interpretation (Shreve, Lacruz, & Angelone, 2010).

*Skip rate* in this study is defined as the percentage of AOIs that are not fixated. In reading research, a word is skipped when it does not receive a direct fixation during first pass (Rayner, Slattery, & Drieghe, 2011). In translation studies, skip rate concerns whether a word has been fixated at all (Schaeffer, Dragsted, Hvelplund, Balling, & Carl, 2016). In this study, skip rate is defined as the percentage of AOIs that do not receive any fixation during the entire trial, because note-reading is a non-linear type of reading (see Section 6.1) and AOIs that are initially skipped have a high probability of being fixated later. In reading research, the skip rate of content words is about 15%, whereas the skip rate of function words is about 65% (Rayner, 1998), leading to an average skipping rate of about one third (Rayner et al., 2011).

Eye movement measures are usually used in combination to gauge cognitive efforts in language processing. For example, if an effect is found on first fixation duration and/or first-pass dwell time, the difficulty was usually experienced immediately on processing that AOI; if an effect is only observed for total and/or second-pass dwell time, this could be indicating a relatively late effect on processing (Liversedge et al., 1998). Examining a series of eye movement measures could provide researchers with abundant information in explaining the cognitive processes during language tasks.

## 6.4 Purpose of the study and research questions

This study attempts to contribute further empirical data to the process research on CI and note-taking using eye tracking. In particular, it reports new data on Phase II where interpreters read their notes and produce a target speech, and seeks answers to the following research questions.

First, how do interpreters read back their notes? This study will be among the first to visualise note-reading in CI. Moreover, when people read, their eye movements differ as a function of the task, such as reading for comprehension, reading for translation and note-reading in CI. This study tries to find out how note-reading resembles and deviates from other forms of reading. To address this research question, the eye tracking measures of average fixation duration, regression rate, and skip rate are consulted.

Second, what is the relationship between the note-taking choices and cognitive load in Phase II of CI? Interpreters make choices on the form and language to be used for note-taking in Phase I of CI, and these choices are likely to impact the cognitive load in Phase II. In Phase I, differences in cognitive load have been observed to be associated with different note-taking choices (Chen, 2017b, 2017c). This study will reveal whether the same patterns of differences will be found in Phase II. A series of eye tracking measures are used to answer this research question, including the first fixation duration, first-pass dwell time, second-pass dwell time, total dwell time, average fixation duration, number of fixations, number of revisits, and skip rate.

## 6.5 Method

### 6.5.1 Participants

An advertisement of the study was sent out via the researchers' professional network to recruit qualified participants (in terms of accreditation, working language combination and experience). A total of 26 professional interpreters participated in the study and they were paid for their participation. They all had the Professional Interpreter accreditation from Australia's National Accreditation Authority for Translators and Interpreters (NAATI), with a working language combination of Mandarin Chinese (native language) and English (non-native language). Most of them had a postgraduate degree in interpreting (65%); some had an interpreting diploma (15%); some attended an intensive interpreting training course (15%); and one was self-trained (4%)[3]. With an average age of 36.4, the participants had worked as full-time or part-time interpreters for an average of 7.4 years. The country they most frequently worked in was Australia (with only two exceptions who worked more frequently in China). For those who were working as part-time interpreters, their other job(s) had a bilingual feature (e.g., interpreter trainer).

The participants used a digital pen and a tablet for note-taking[4] and their eye movements were recorded using a head-mounted eye tracker (there was a practice session to familiarise them with the equipment, see Section 5.4). A post-experiment

[3] Percentages do not add up to 100% due to rounding.

[4] The pen recording data has been reported in another article currently under review.

questionnaire was used to ask whether the interpreters felt comfortable with the digital pen and tablet. If the participants indicated a high level of discomfort with the equipment (measured as a rating below 50% on the comfort scale), the participant was excluded from the final data analysis due to the high likelihood that the use of the equipment would be too much of a barrier to that participant to allow for ecologically valid data. Eye tracking also led to some further data loss (e.g., the eye tracker did not work well with bi-focal glasses), which was not uncommon in eye tracking studies. In the end, 18 participants had both their pen and eye data successfully collected, meaning that roughly 31% of the participants were excluded. The data loss in eye tracking research varies significantly, ranging from 20–60% of participants/trials (Holmqvist et al., 2011). Considering that the data loss in this study is a result of the combination of two methods (pen recording and eye tracking), it is within an acceptable range.

### 6.5.2 Apparatus

The eye tracker used in the study was the *SMI ETG 2W*. It is a lightweight (47 g), head-mounted eye tracker that uses dark pupil tracking. It has a tracking accuracy of .5° over all distances and a sampling rate of 60 Hz. The eye tracker has a built-in high-definition camera for scene recording. Participants sat in front of a 13-inch LCD tablet with a resolution of 1366×768 pixels. The tablet was the *Wacom Cintiq 13HD*, equipped with a *Wacom Pro Pen*, which the participants used for note-taking. The *Eye and Pen* software[5] was used for collecting and analysing the pen data. The experiment took place in a sound-proof studio with constant artificial illumination. The SMI software *iView ETG* and *BeGaze* were used with default settings for eye data recording and analysis respectively.

### 6.5.3 Tasks

There were two CI tasks, one from English to Chinese (E-C), and one from Chinese to English (C-E). A series of steps were taken to ensure that the two tasks were as comparable as possible.

[5] More introduction to the software can be found on http://www.eyeandpen.net/?lng=en. The software can be used for the synchronous recording of eye and pen data, but unfortunately it does not support the type of eye tracker used in this study. So all the eye data was synchronised with the pen data post-hoc. This added more work load to the researcher, but did not affect data quality.

*Step one, source text selection and editing*. Two English video clips on similar topics (see Table 6.2) were selected from the Internet and transcribed by the author. This created two English scripts (1 and 2) which were later edited to make them comparable by an Australian university lecturer. The focus was on the length, complexity and style of language. The edited scripts were put into CPIDR 5[6] for analysis and results showed that they were comparable in word count, proposition count and idea density (Table 6.2).

*Step two, translation and editing*. Script 2 was translated into Chinese and refined to make it suitable to be read aloud and recorded as a speech. This was done by two Chinese editors working at a local Chinese radio station.

*Step three, audio recording*. A native English speaker in Australia[7] (the same person who edited the original English scripts) read aloud the English speech and the audio was recorded. A native Chinese (Mandarin) speaker (a radio personality from the radio station mentioned above) read aloud the Chinese speech and the audio was recorded. Both recordings were made in professional sound-proof studios.

*Step four, audio editing*. The recordings were edited using Audacity. Each speech was divided into three segments and controlled for variables such as pauses, duration and speed (Table 6.2).

Table 6.2 Task specifications

| Task | Topic | Word/character count | Proposition count | Idea density | Duration | Segment duration | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 1 | 2 | 3 |
| E-C | How to register a business in Australia | 631 | 321 | 0.509 | 4'59" | 1'18" | 2'02" | 1'39" |
| C-E | How to purchase a property in Australia | 630 (English) * 944 (Chinese) | 324* | 0.514* | 4'47" | 1'10" | 2'07" | 1'30" |

* The calculations were based on the original English script to enable comparison between the two tasks.

### 6.5.4 Experimental procedure

First, participants were allowed sufficient time to practice using a short task. During this practice session they became familiarised with the digital pen and tablet, the eye tracker, and the experiment procedures. Second, the participants performed the two interpreting

[6] More details about the programme and the analysis can be found at http://ai1.ai.uga.edu/caspr/ and Brown, Snodgrass, Kemper, Herman, and Covington (2008).

[7] The person speaks standard English with a neutral accent.

tasks with randomised order. Third, after both tasks were completed, the participants were provided with their written notes for cued retrospection. Fourth, the participants completed a post-experiment questionnaire. The purpose of the questionnaire was to collect information such as the participants' familiarity with the task topics, how they felt about using the digital pen and the eye tracker, and other feedback about the experiment. A calibration[8] procedure was carried out for each participant at the beginning of each trial. It was followed by a validation procedure and recalibration was performed when needed. The calibration was further checked at the end of each trial. The experimental procedure is outlined in Table 6.3.

Table 6.3 Experimental procedure

| Session | | Steps |
|---|---|---|
| I | Practice | Eye tracker set up, calibration, and start a trial;<br>Start the Eye and Pen software;<br>Let the participant practice note-taking with the digital pen and tablet using the practice material, until they feel comfortable with the equipment and familiar with the procedures. |
| II | Task 1 | Calibration, validation (recalibration if needed), and start a trial;<br>Start the Eye and Pen software;<br>Perform Task 1 (could be either E-C or C-E, randomised);<br>Recheck the calibration;<br>Rest. |
| | Task 2 | Same procedures as Task 1. |
| III | Retrospection | Retrospection for Task 1 and Task 2. |
| IV | Questionnaire | Complete a post-experiment questionnaire. |

### 6.5.5 Data and analysis

#### 6.5.5.1 Semantic Gaze Mapping and AOI drawing

Semantic Gaze Mapping is a function provided by SMI's analysis software BeGaze. This function is used to map the gaze data points from scene videos to corresponding reference images. There are two ways to perform the mapping: an event-based mapping where there is one gaze mapping for each event at a certain frame and the other frame mappings for that event are generated automatically; a frame-based mapping where

[8] The SMI ETG offers two types of calibration: a one-point and a three-point calibration. The one-point calibration was used in this study because it generated good quality data in test runs and it reduced the time and money cost of the experiment, during which professional interpreters were paid for their participation.

there is one gaze mapping for each frame in the scene video. The latter is very labour-intensive and is usually used when multiple analysers are available to do the mapping. In this study, there was only one researcher to do the mapping so an event-based mapping method was selected.

Each page of notes taken by each interpreter was saved automatically by the Eye and Pen software as a picture file. These pictures were imported into BeGaze to be used as reference images. The gaze point data on the scene video were mapped onto the reference images. Figure 6.1 shows two screenshots during Semantic Gaze Mapping. On the right is the scene video of one participant reading one page of notes, showing the current gaze position. On the left is a corresponding reference image (a picture file of the page of notes being read). When the analyser holds the mouse button while clicking the correct gaze position on the reference image, a magnified image of the area under the mouse cursor is shown for improved positioning.

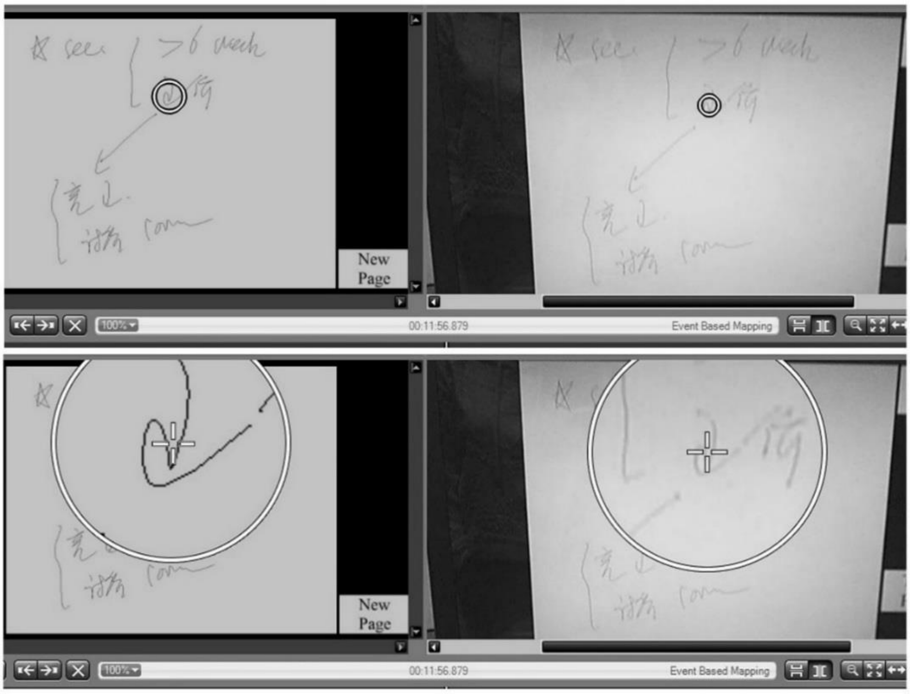


Figure 6.1 An example of Semantic Gaze Mapping

After all relevant eye tracking data were mapped onto the reference images, AOIs were drawn on the images instead of the scene video, which increased the accuracy and efficiency of analysis. An AOI was drawn for each note unit (Figure 6.2).

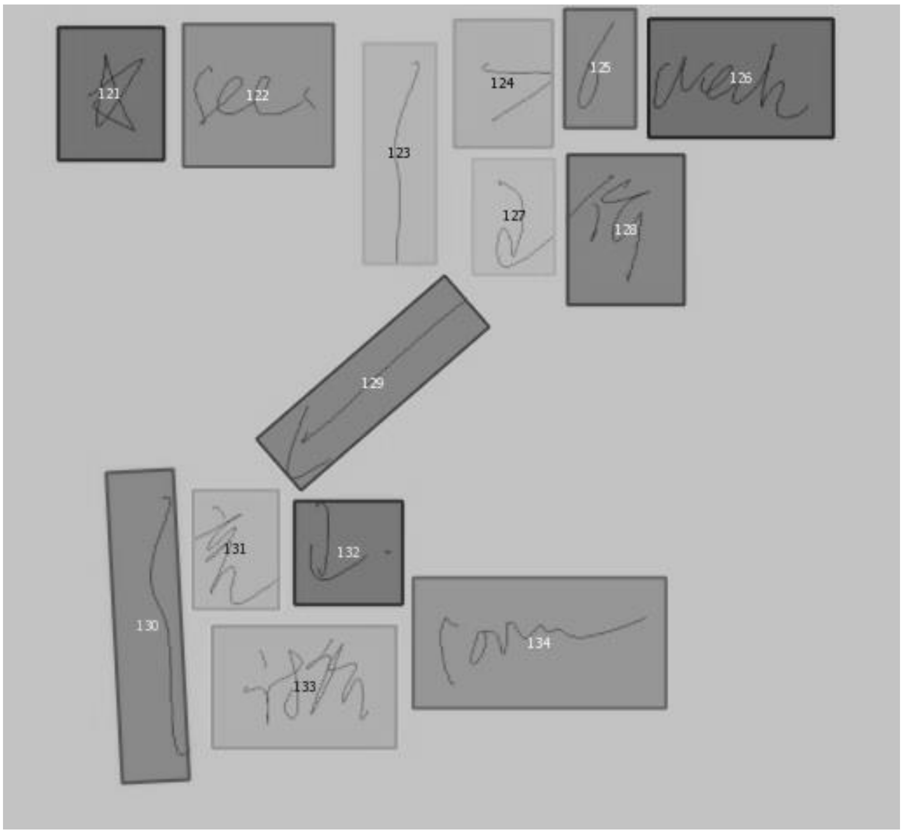


Figure 6.2 An example of drawing the AOIs

This study compares how interpreters read notes in different forms (language vs. symbol notes; abbreviation vs. full word notes) and in different languages (Chinese vs. English notes). Therefore, each AOI was labelled post hoc according to its form and language (Figure 6.3). The labelling of the notes was informed by the cued retrospection, during which participants were asked to provide as much information as they could remember about the note-taking process as well as the form and language of each note unit. This was important because interpreters' notes were highly individualised and could be difficult for others to decipher. One note unit could have multiple labels. For example, an English abbreviation could be labelled simultaneously as "language", "abbreviation" and "English". AOIs with the same label are treated as one single (distributed) AOI. For example, all notes labelled with "Chinese" belong to an AOI called Chinese and all notes labelled with "English" belong to an AOI called English.

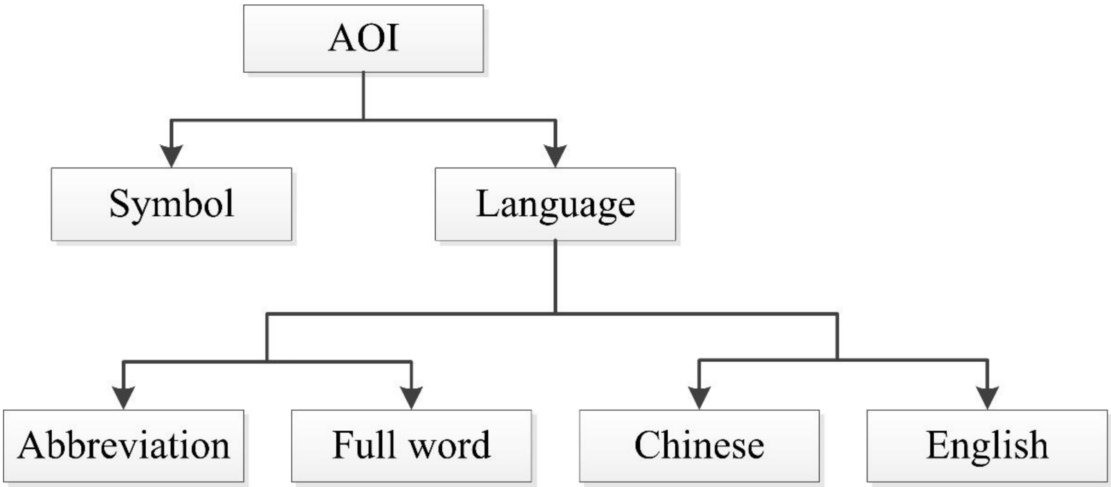


Figure 6.3 Labelling the AOIs

#### 6.5.5.2 Eye tracking measures used in this study

A range of eye tracking measures (Table 6.4) has been used in this study to answer the two research questions. These measures are summarised with a brief definition provided.

Table 6.4 Eye tracking measures used in this study

| Eye tracking measure | Definition | Target research question |
|---|---|---|
| Regression rate | Number of regressive fixations to previous AOIs divided by the total number of fixations. | How do interpreter read back their notes? |
| First fixation duration | Duration of the first fixation on an AOI. | What is the relationship between the note-taking choices and cognitive load in Phase II of CI? |
| First-pass dwell time | Sum of the fixation durations during the first dwell on an AOI from entry to exit. | |
| Second-pass dwell time | Total dwell time – first-pass dwell time. | |
| Total dwell time | Sum of the fixation durations on an AOI in an entire trial. | |
| Number of fixations | Fixation count inside an AOI during an entire trial. | |
| Number of revisits | Count of transitions to an AOI already visited. | |
| Average fixation duration | Total dwell time / number of fixations | Both research questions |
| Skip rate | Number of AOIs skipped divided by the total number of AOIs. | |

#### 6.5.5.3 Statistical analysis

The unit of analysis in this study was a note unit and all data were standardised in the same manner. For example, if the number of Chinese notes written by a participant is $n$, then the total dwell time (T) of Chinese notes of that participant is calculated as:

$$\frac{1}{n} \times \sum_{i=1}^{n} T_i$$

Paired-samples t-tests were conducted to compare the note-taking choices in different forms (language vs. symbol; abbreviation vs. full word) and languages (Chinese vs. English). All statistical analyses were performed by running the IBM SPSS Statistics 22. Two-tailed $p$ values less than 0.05 were considered to be statistically significant. Cohen's $d$ (the difference between the means divided by the pooled standard deviation) was used to indicate the effect sizes, which were classified as small ($d$ = 0.2), medium ($d$ = 0.5), and large ($d$ = 0.8).

Normality was checked for all data using the Shapiro-Wilk (S-W) test. Of the 96 sets of data involved in t-tests (8 eye movement measures × 2 interpreting directions × 6 note-taking choices), only three were not normally distributed, all found in the C-E task: the total dwell time of Chinese notes ($SW$ = .87, $df$ = 18, $p$ = .015, skewness = 1.09, kurtosis = 0.24), second-pass dwell time of Chinese notes ($SW$ = .87, $df$ = 18, $p$ = .017, skewness = 1.41, kurtosis = 1.82), and the first-pass dwell time of full word notes ($SW$ = .79, $df$ = 18, $p$ = .001, skewness = 2.14, kurtosis = 6.60). For comparisons involving these three sets of data, additional non-parametric tests (the related-samples Wilcoxon signed-rank test) were run and the results (non-significant) did not change. Considering that parametric statistics are robust to violations of the normality assumption (Tabachnick & Fidell, 2007), and for the sake of consistency in data report and interpretation, only the t test results were presented in this study.

## 6.6 Results

### 6.6.1 How interpreters read back their notes

A visual check of the scanpath video revealed that note-reading progressed in a non-linear manner. A sample AOI sequence chart is presented in Figure 6.4. This shows how one participant read one page of notes. The x axis indicates the passing of time and the y axis shows the AOIs (notes) being fixated (the little bars are fixations). If note-reading was linear, the participant would read the notes one by one in a sequential manner as time passed. However, the data showed that notes were read in groups, with each group of notes being meaning-related chunks. The participant read the first group of notes (AOIs 17 to 20) between time 0 to 8750 milliseconds, moved to the second group of notes (AOIs 22 to 26) and read them between time 8750 and 21,250 milliseconds, and then read the third group of notes (AOIs 28 to 33) for the remaining time. The participant also moved frequently between the notes within each group, indicating that during speech production around one note unit, the interpreter did read ahead to examine other related note units. But this read-ahead usually happened within the note groups (meaning-based chunks) and rarely between the groups.

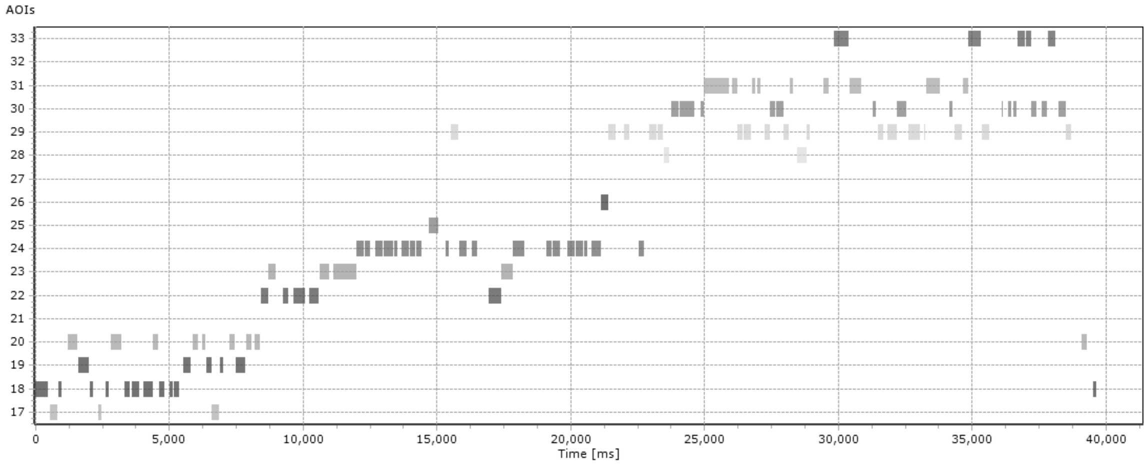


Figure 6.4 A sample AOI sequence chart

Descriptive statistics of the eye movement measures during note-reading are reported in Table 6.5. The average fixation duration was around 277 milliseconds, the regression rate was about 23% and the skip rate was about 12% averaged across the two tasks. All note units were included in the analysis regardless of their form and language.

Table 6.5 Descriptive statistics of note-reading

| | E-C | | C-E | | Average | |
|---|---|---|---|---|---|---|
| Measure | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Average fixation duration (millisecond) | 271.48 | 70.99 | 281.77 | 76.34 | 276.63 | 72.93 |
| Regression rate (percentage) | 23.18 | 2.91 | 21.86 | 3.00 | 22.52 | 2.73 |
| Skip rate (percentage) | 11.81 | 6.03 | 11.40 | 5.52 | 11.61 | 5.46 |

### 6.6.2 The relationship between the note-taking choices and cognitive load in Phase II of CI

#### 6.6.2.1 The choice between language and symbol

Significant differences were found between language and symbol notes in all the measures (Table 6.6). The total, first-pass and second-pass dwell times, the first fixation duration, and the average fixation duration of language notes were significantly longer than those of symbol notes. The number of fixations and revisits of language notes were significantly higher than those of the symbol notes, whereas the skip rate of the former was significantly lower than that of the latter. Further, Cohen's effect size values (Cohen's *d*, *M* = 2.53) suggested large magnitude of these differences, much bigger than the differences found between abbreviation and full word (6.2.2) and between Chinese and English (6.2.3).

Table 6.6 The choice between language and symbol

| Task | Measure | Language | | Symbol | | | | |
|---|---|---|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | First fixation duration | 260.03 | 61.13 | 141.44 | 53.81 | 13.00*** | 17 | 3.07 |
| | First-pass dwell time | 369.82 | 86.68 | 165.39 | 71.63 | 15.80*** | 17 | 3.72 |
| | Second-pass dwell time | 1261.76 | 576.24 | 288.38 | 163.76 | 8.67*** | 17 | 2.04 |
| | Total dwell time | 1631.57 | 626.20 | 453.77 | 212.86 | 9.83*** | 17 | 2.32 |
| | Avg fixation duration | 266.76 | 73.89 | 151.12 | 61.09 | 9.99*** | 17 | 2.35 |
| | Number of fixations | 5.98 | 1.97 | 1.73 | 0.74 | 10.58*** | 17 | 2.49 |
| | Number of revisits | 2.72 | 1.06 | 0.79 | 0.46 | 10.89*** | 17 | 2.57 |
| | Skip rate | 0.05 | 0.05 | 0.41 | 0.13 | -12.93*** | 17 | 3.05 |
| C-E | First fixation duration | 276.65 | 78.96 | 153.56 | 68.08 | 10.80*** | 17 | 2.55 |
| | First-pass dwell time | 418.49 | 132.10 | 190.46 | 81.63 | 8.28*** | 17 | 1.95 |
| | Second-pass dwell time | 1265.83 | 530.43 | 280.11 | 165.94 | 8.88*** | 17 | 2.09 |
| | Total dwell time | 1684.32 | 591.79 | 470.57 | 226.51 | 9.76*** | 17 | 2.30 |
| | Avg fixation duration | 282.51 | 82.48 | 158.59 | 63.27 | 10.01*** | 17 | 2.36 |
| | Number of fixations | 6.06 | 1.75 | 1.87 | 0.89 | 11.73*** | 17 | 2.77 |
| | Number of revisits | 2.60 | 0.86 | 0.87 | 0.54 | 12.45*** | 17 | 2.93 |
| | Skip rate | 0.04 | 0.03 | 0.37 | 0.18 | -8.11*** | 17 | 1.91 |

Notes:
1. *** $p < .001$;
2. All time measures were calculated in milliseconds;
3. Avg = average.

#### 6.6.2.2 The choice between abbreviation and full word

Data on the choice between abbreviation and full word notes are summarised in Table 6.7. In both directions of interpreting, significant differences were found between abbreviations and full words in terms of the total and second-pass dwell times, the number of fixations and the number of revisits (the number of revisits in the C-E task was only approaching significance, $p = .051$). Full words had longer dwell times, more fixations and more revisits than abbreviations. The skip rate of abbreviations was significantly higher than that of full words in the E-C task but not in the C-E task.

Cohen's d values ($M$ = 0.81) suggested that most of the effect sizes were large. In neither direction were the first-pass dwell time, first fixation duration and average fixation duration significantly different.

Table 6.7 The choice between abbreviation and full word

| | | Abbreviation | | Full word | | | | |
|---|---|---|---|---|---|---|---|---|
| Task | Measure | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | First fixation duration | 264.37 | 65.26 | 256.57 | 63.49 | 1.04 | 17 | |
| | First-pass dwell time | 370.70 | 99.24 | 378.77 | 85.07 | -0.55 | 17 | |
| | Second-pass dwell time | 1106.25 | 536.17 | 1401.71 | 632.42 | -3.19** | 17 | 0.75 |
| | Total dwell time | 1476.96 | 578.05 | 1780.48 | 678.39 | -3.51** | 17 | 0.83 |
| | Avg fixation duration | 268.87 | 76.30 | 263.96 | 70.94 | 1.17 | 17 | |
| | Number of fixations | 5.37 | 1.98 | 6.62 | 2.02 | -4.54*** | 17 | 1.07 |
| | Number of revisits | 2.52 | 1.11 | 2.94 | 1.02 | -3.55** | 17 | 0.84 |
| | Skip rate | 0.06 | 0.06 | 0.04 | 0.04 | 3.66** | 17 | 0.86 |
| C-E | First fixation duration | 278.46 | 84.33 | 271.96 | 73.56 | 0.57 | 17 | |
| | First-pass dwell time | 409.75 | 120.55 | 434.10 | 191.17 | -0.62 | 17 | |
| | Second-pass dwell time | 1149.45 | 393.43 | 1374.86 | 635.71 | -2.88** | 17 | 0.68 |
| | Total dwell time | 1559.19 | 436.88 | 1808.96 | 731.32 | -2.76* | 17 | 0.65 |
| | Avg fixation duration | 284.40 | 87.56 | 277.59 | 75.26 | 0.70 | 17 | |
| | Number of fixations | 5.60 | 1.38 | 6.58 | 2.21 | -3.37** | 17 | 0.80 |
| | Number of revisits[4] | 2.48 | 0.69 | 2.73 | 1.07 | -2.10 | 17 | |
| | Skip rate | 0.05 | 0.04 | 0.04 | 0.05 | 1.26 | 17 | |

Notes:
1. * $p < .05$, ** $p < .01$, *** $p < .001$;
2. All time measures were calculated in milliseconds.
3. Std = standardised, avg = average.
4. The number of revisits in the C-E task was closely significant ($p$ = .051, $d$ = 0.49).

Considering that there might be a consistent difference between the length of abbreviation and full word notes, and that this difference might have caused the differences observed in some of the measures, this study took a further step to control for length. The notes were grouped according to the number of characters they contain (e.g., one-character group, two-character group, etc.). Then, notes in each group were compared. When abbreviations and full words with the same length were compared with each other, the differences were no longer significant. That is to say, full word notes were not looked at more than abbreviation notes when the variable of length was controlled.

#### 6.6.2.3 The choice between Chinese and English

Data on the choice between Chinese and English notes are summarised in Table 6.8. When the direction of interpreting was English to Chinese, the total and second-pass dwell time of Chinese notes were significantly shorter than those of English notes; the average fixation duration of Chinese notes were shorter than those of English notes (approaching significance, $p$ = .058); the number of fixations and revisits on the Chinese notes were significantly lower than those on the English notes; and the Chinese notes had a significantly higher skip rate. Notes in the two languages did not differ in the first-pass dwell time or the first fixation duration. However, when the direction of interpreting was Chinese to English, no significant difference was found on any of the measures between the two languages.

Table 6.8 The choice between Chinese and English

| | | Chinese | | English | | | | |
|---|---|---|---|---|---|---|---|---|
| Task | Measure | *M* | *SD* | *M* | *SD* | *t* | *df* | *d* |
| E-C | First fixation duration | 250.96 | 81.27 | 261.02 | 63.69 | -0.81 | 15 | |
| | First-pass dwell time | 354.56 | 152.25 | 381.96 | 90.41 | -0.92 | 15 | |
| | Second-pass dwell time | 983.18 | 543.39 | 1284.23 | 492.54 | -7.00*** | 15 | 1.75 |
| | Total dwell time | 1337.74 | 631.96 | 1666.19 | 563.95 | -6.20*** | 15 | 1.55 |
| | Avg fixation duration[5] | 255.61 | 84.02 | 271.28 | 77.08 | -2.06 | 17 | 0.51 |
| | Number of fixations | 4.86 | 2.22 | 6.10 | 1.70 | -4.70*** | 15 | 1.17 |
| | Number of revisits | 2.24 | 0.95 | 2.70 | 0.75 | -3.54** | 15 | 0.89 |
| | Skip rate | 0.10 | 0.13 | 0.04 | 0.05 | 2.33* | 15 | 0.58 |
| C-E | First fixation duration | 277.38 | 93.57 | 271.41 | 82.71 | 0.46 | 17 | |
| | First-pass dwell time | 414.65 | 151.36 | 405.75 | 141.55 | 0.39 | 17 | |
| | Second-pass dwell time | 1256.62 | 588.73 | 1246.50 | 519.32 | 0.20 | 17 | |
| | Total dwell time | 1671.27 | 606.34 | 1652.24 | 597.69 | 0.32 | 17 | |
| | Avg fixation duration | 294.04 | 93.70 | 276.63 | 87.04 | 1.79 | 17 | |
| | Number of fixations | 5.79 | 1.59 | 6.07 | 1.80 | -1.11 | 17 | |
| | Number of revisits | 2.52 | 0.82 | 2.55 | 0.84 | -0.23 | 17 | |
| | Skip rate | 0.05 | 0.04 | 0.04 | 0.04 | 0.58 | 17 | |

Notes:

1. * $p < .05$, ** $p < .01$, *** $p < .001$;
2. All time measures were calculated in milliseconds;
3. The degree of freedom (df) in the E-C task was 15 because two interpreters took notes in English only;
4. Std = standardised, avg = average.
5. The average fixation duration in the E-C task was closely significant ($p$ = .058, $d$ = 0.51).

## 6.7 Discussion

### 6.7.1 Note-reading, reading for comprehension and reading for translation

Qualitative data such as scanpaths and AOI sequence charts confirm that note-reading proceeds in a non-linear manner. Instead of reading the notes one by one, interpreters move from one group of notes to another. Each group of notes are possibly meaning related chunks, corresponding to one part of the target speech. Reading-ahead happens frequently within each note group, but interpreters do not usually read ahead to examine another group of notes until they complete reading the current note group.

Quantitative data shows that note-reading is a unique reading behaviour, differing from other forms of reading in many ways. The average fixation duration in note-reading is found to be 277 milliseconds. In reading research, the reported average fixation duration is 225 milliseconds in silent reading and 275 milliseconds in oral reading (Rayner, 1998). This seems to indicate that the reading of notes corresponds with oral reading. In translation research, the reported average fixation durations tend to vary between studies: from 205 milliseconds (Jakobsen & Jensen, 2008) to 245 milliseconds (Dragsted, 2010) in reading silently (in preparation for translation); from 235 milliseconds (Jakobsen & Jensen, 2008) to 252 milliseconds in sight translation/interpretation (Dragsted & Hansen, 2009).

The regression rate in note-reading found in this study is 23%, almost twice as much as the regression rate reported in reading research, which is 10–15% (Rayner, 1998). It is similar to the regression rate found in reading in a foreign language (20–25%) but lower than that in sight translation/interpretation (30–35%) (Shreve et al., 2010).

The skip rate in note-reading is found to be 12% across all note forms. Compared to the average skip rate of about one third in reading (Rayner et al., 2011), the skip rate in note-reading is much lower. This makes sense due to the high meaning load carried by note units.

Considering that a longer average fixation duration, higher regression rate, and lower skip rate usually indicate higher cognitive load (see Section 3), the data seem to suggest that the level of cognitive load during note-reading is higher than that during reading silently (both reading for comprehension and reading in preparation for translation) and more similar to that during reading in sight translation/interpretation. This is plausible considering that when interpreters read their notes, they are simultaneously involved with the additional tasks of memory retrieval, translation and speech production.

### 6.7.2 Note-taking choices and cognitive load during Phase II of CI

Highly significant results ($p < .001$) with large effect sizes (Cohen's $d$, $M = 2.53$) were found between language and symbol notes, suggesting that language notes are much more cognitively demanding to read than symbol notes. Notes in these two forms differ not only in early and fast cognitive processing such as recognising and identifying the note unit and its meaning (as indicated by first fixation duration and first-pass dwell time), but also in slower operations such as linking the note unit to the context and producing a target speech around it (as indicated by second-pass dwell time and total dwell time). Furthermore, eye tracking measures such as average fixation duration, number of revisits, and skip rate seem to suggest that reading language and reading symbol could almost be treated as different task types. Their differences are even bigger than those between different types of reading tasks (e.g., reading for comprehension and reading for translation).

In terms of the choice between abbreviation and full word notes, significant differences are observed in second-pass and total dwell time, but not in first fixation duration or first-pass dwell time. The two note forms differ in number of fixations and revisits, but are similar in average fixation durations. That is to say, full word notes are not more difficult to recognise or comprehend than abbreviations, but they receive more fixations and revisits, which lead to longer second-pass and total dwell times. The observed differences could be a function of length: it is possible that full words are simply being looked at more than abbreviations because they are longer. A further step needs to be taken to examine the difference between these two forms while controlling for length.

The notes were grouped according to the number of characters they contain. When notes in each group were compared (i.e., abbreviations and full words with the same length), the differences were no longer significant. That is to say, full words are looked at more mainly because they are longer than abbreviations. However, considering that time is a valuable resource in interpreting and that full words do cost more time during note-reading, abbreviations might be a more favourable choice because they attract less visual attention and take less time to process.

The observed difference between Chinese and English notes is only significant in the English to Chinese direction. Notes in the two languages have similar first fixation durations and first-pass dwell times, indicating that the levels of cognitive load associated with initial processes such as identifying and recognising notes are similar. However, Chinese notes have shorter average fixation durations, second-pass and total dwell times and receive fewer fixations and revisits. This indicates that Chinese notes are easier to process when it comes to later cognitive processes such as integrating the note unit with the context and producing a target speech around it.

There are two possible contributing factors. First, Chinese is the native language of the interpreters, hence reading Chinese notes (native language processing) is likely to be less demanding than reading English notes (non-native language processing). Second, note-reading and speech production in the same language (within-language processing) is likely to be less cognitively demanding than reading one language and speaking in another language (between-language processing). Therefore, when interpreting from English to Chinese, note-reading in Chinese is both native-language processing and within-language processing, leading to a significantly lower level of cognitive load associated with Chinese notes. However, when the direction of interpreting is reversed, the effect is diminished by the extra costs brought by between-language processing (from Chinese notes to English speech).

As has been mentioned in Section 2 (Table 6.1), the varied note-taking choices made by interpreters are associated with different levels of cognitive load in Phase I of CI. Now it is time to put the two pieces of the puzzle together and see how note-taking choices and cognitive load interplay during the entire process of CI (Table 6.9). Taken together, the evidence suggests that cognitive demand (rather than physical or temporal demands) is the main determining factor in interpreters' note-taking choices. Interpreters would take proactive actions to reduce cognitive load during Phase I of CI, leaving (either consciously or unconsciously) the extra burden to Phase II. However, the note-taking choices made by interpreters with the intention to ease the cognitive system in Phase I are not always optimal when all factors are taken into account. For example, interpreters would benefit more from using more symbols considering that they can save effort both in note-writing and note-reading. They could also use more abbreviations, especially considering that abbreviations save time and that they are not more difficult to process than full words.

Table 6.9 Physical, temporal and cognitive demands of different note-taking choices in both Phase I and II of CI

| | Choice of form | | Choice of language |
|---|---|---|---|
| | Language vs. symbol | Abbreviation vs. full word | Chinese vs. English |
| Physical and temporal demands | Language > Symbol | Abbreviation < Full word | Chinese ≈ English |
| Cognitive load in Phase I | Language < Symbol | Abbreviation ≈ Full word | Chinese > English (only significant in the C-E task) |
| Preferred note-taking choice | Language | No preference | English |
| Cognitive load in Phase II | Language > Symbol | Abbreviation < Full word | Chinese < English (only significant in the E-C task) |

## 6.8 Conclusions

This study is among the first to reveal the processes that occur during Phase II of CI. It contributes both qualitative and quantitative data to help explain the unique task of note-reading and to present the relationship between note-taking choices and cognitive load in the entire interpreting process. It reveals that there is a trade-off between the cognitive costs associated with the different note-taking choices in the two phases of CI. Generally speaking, note-taking choices that help to reduce cognitive load in Phase I are more difficult to process during Phase II. This study is also pioneering in the use of eye tracking to investigate CI. Eye tracking has proven to be a useful and powerful method for visualising and quantifying the process of note-reading and CI.

This exploratory study is admittedly limited in several ways. First, the only head-mounted eye tracker available to the study was the SMI ETG, with a relatively low sampling rate of 60 Hz. If other types of head-mounted eye trackers with higher sampling rates are available, more high quality data can be collected and more eye movement measures (such as saccades and velocity-based measures) could be used for analysis. Second, the inferences in the study are made from data successfully collected from a small group of 17 professional interpreters. Given the common issue of data loss in eye tracking research (see Section 5.1), the study could have benefited from a larger sample size. Third, the comparisons between note-reading and other types of reading are being made even though the stimulus used are possibly very different. The study has been careful not to over-interpret the data because the difficulty of the stimulus materials might contribute to the differences observed in data. Fourth, it is not sure whether the findings of this study can be generalised to other language pairs or to other types of participants (e.g., student interpreters). Hopefully these limitations can provide some ideas for future research to be conducted on related topics.

Cognitive and process-oriented investigations on the task of interpreting is an attractive field as it could reveal much about the way in which the human mind processes two languages simultaneously. There is much that remains unknown, opening the door for many future studies. The findings of this study will not only benefit the field of interpreting itself, but also contribute to our understanding of language processing and the human mind and articulate further links between disciplines.

## 6.9 References


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting, 14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Andres, D. (2002). *Konsekutivdolmetschen und notation [Consecutive interpreting and note-taking]*. Frankfurt: Peter Lang.

Andres, D. (2015). Consecutive interpreting. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 84–87). Abingdon: Routledge.

Bowen, D., & Bowen, M. (1980). *Steps to consecutive interpretation*. Washington, DC: Pen and Booth.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40*(2), 540–545. doi:10.3758/BRM.40.2.540

Cardoen, H. (2013). The effect of note-taking on target-text fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging research in translation studies: Selected papers of the CETRA Research Summer School 2012* (pp. 1–22). Leuven: CETRA.

Carl, M., & Kay, M. (2011). Gazing and typing activities during translation: A comparative study of translation units of professional and student translators. *Meta, 56*(4), 952–975. doi:10.7202/1011262ar

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

Chen, S. (2017a). The construct of cognitive load in interpreting and its measurement. *Perspectives*. doi:10.1080/0907676X.2016.1278026

Chen, S. (2017b). Note-taking in consecutive interpreting: New data from pen recording. *Translation and Interpreting*, *9*(1), 4–23. doi:10.12807/ti.109201.2017.a02

Chen, S. (2017c). *The process of note-taking and consecutive interpreting: Evidence from digital pen recording*. Manuscript submitted for publication.

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research*, *39*(2), 136–144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting, 6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters' notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V. (2010). Consecutive interpreting. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies: Volume 1* (pp. 75–79). Amsterdam: John Benjamins.

Daneman, M., Reingold, E. M., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 884–898.

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

Dragsted, B. (2010). Coordination of reading and writing processes in translation: An eye on uncharted territory. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 41–62). Amsterdam: Jonh Benjamins.

Dragsted, B. (2012). Indicators of difficulty in translation: Correlating product and process data. *Across Languages and Cultures, 13*(1), 81–98. doi:10.1556/Acr.13.2012.1.5

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

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (2 ed.). London: Springer-Verlag.

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. Manchester and Kinderhook: St. Jerome.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Huey, E. B. (1908). *The psychology and pedagogy of reading*. New York: Macmillan.

Hvelplund, K. T. (2017). Eye tracking in translation process research. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 248-264). Hoboken, NJ: John Wiley & Sons.

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

Ilg, G., & Lambert, S. (1996). Teaching consecutive interpreting. *Interpreting, 1*(1), 69-99.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Oxford: Elsevier.

Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 103-122). Frederiksberg: Samfundslitteratur.

Jensen, K. T. H., Sjørup, A. C., & Balling, L. W. (2009). Effects of L1 syntax on L2 translation. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology,* *technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (pp. 319-336). Frederiksberg: Samfundslitteratur.

Jones, R. (1998). *Conference interpreting explained*. Manchester: St. Jerome.

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

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Liversedge, S. P., Gilchrist, I. D., & Everling, S. (Eds.). (2011). *The Oxford handbook of eye movements*. Oxford: Oxford University Press.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford: Elsevier.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology, 4*(4), 565–572. doi:10.1037/0096-1523.4.4.565

Matyssek, H. (1989). *Handbuch der notizentechnik für dolmetscher [Handbook of note-taking for interpreters]*. Heidelberg: Julius Groos.

Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., . . . Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology, 12*(7), 830–840. doi:10.1016/j.acra.2005.03.068

Murray, W. S. (2000). Sentence processing: Issues and measures. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 649–664). Oxford: Elsevier.

Orlando, M. (2010). Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter, 15*, 71–86.

Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation and Interpreting, 6*(2), 39–54.

Pollatsek, A., & Rayner, K. (1990). Eye movements and lexical access in reading. In D. A. Balota, G. B. F. d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 143–164). Hillsdale, NJ: Erlbaum.

Pynte, J. (1996). Lexical control of within-word eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 958–969. doi:10.1037/0096-1523.22.4.958

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4*(3-4), SI21–SI49. doi:10.1080/01690968908406362

Rayner, K., Slattery, T. J., & Drieghe, D. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 514–528.

Rozan, J.-F. o. (1956). *La prise de notes en interprétation consécutive [Note-taking in consecutive interpreting]*. Genève: Georg.

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., & Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB* (pp. 183–210). Cham: Springer.

Schmauder, A. R. (1992). Eye movements and reading processes. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 369–378). New York: Springer-Verlag.

Seleskovitch, D., & Lederer, M. (1995). *A systematic approach to teaching interpretation* (J. Harmer, Trans.). Silver Spring, MD: Registry of Interpreters for the Deaf.

Setton, R., & Dawrant, A. (2016). *Conference interpreting: A trainer's guide*. Amsterdam: John Benjamins.

Sharmin, S., Špakov, O., Räihä, K.-J., & Jakobsen, A. L. (2008). Where on the screen do translation students look while translating, and for how long? In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 31–51). Frederiksberg: Samfundslitteratur.

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

Sjørup, A. C. (2008). Metaphor comprehension in translation: Methodological issues in a pilot study. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (pp. 53–77). Frederiksberg: Samfundslitteratur.

Szabó, C. (2006). Language choice in note-taking for consecutive interpreting. *Interpreting, 8*(2), 129–147. doi:10.1075/intp.8.2.02sza

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, Ł., & Kruger, J.-L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures, 17*(2), 183–204. doi:10.1556/084.2016.17.2.3

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th edition)*. Boston: Pearson.

# Chapter 7 Summary, discussion and conclusion

From a cognitive perspective, CI is a remarkable language processing task because it involves multiple tasks in both phases: listening, analysing, note-taking and memorising in Phase I, and remembering, note-reading, and producing a translated speech in Phase II. Note-taking is a unique activity that channels the two phases of CI, leaving traces of the cognitive processes involved in the interpreting process. This PhD study has approached note-taking and CI from a cognitive perspective, with particular focus on cognitive load in the process. This concluding chapter summarises the main findings of the PhD research and discusses their implications, pinpoints the strengths and limitations of the study, and proposes some promising areas for future research.

## 7.1 Main findings of the research

Three main questions concerning note-taking and CI have been addressed in this study:

(1) What are the preferred note-taking choices of professional interpreters in terms of form and language?

(2) What is the relationship between note-taking choices and cognitive load in CI?

(3) What is the relationship between note-taking and interpreting performance?

In this section, the main findings of the research are elaborated and organised around the above questions.

### 7.1.1 Note-taking preferences

When interpreters take notes, they make choices on the form of notes: whether to use language or symbol, and if language, whether to use abbreviation or full word. Literature suggests that there is a consistent preference for language over symbol established in both professional interpreters (Andres, 2002; Dam, 2004a) and student interpreters (Andres, 2002; Dam, 2004b; Liu, 2010; Lung, 2003; Wang, Zhou, & Wang, 2010). Findings on the choice between abbreviations and full words, however, appear to be much more inconsistent. While some studies find their participants prefer abbreviations (Dai & Xu, 2007; Wang et al., 2010), others find the opposite (Dam, 2004a; Liu, 2010; Lung, 2003).

In this study (see Chapter 4 and Chapter 5), a clear dominance of language notes over symbol notes was detected in the notes taken by professional interpreters in both directions of interpreting, corroborating the previous studies. The empirical data to date conjointly support a general language-based note-taking system, rather than a symbol-based system. This study did not find a significant difference between the distribution of notes over abbreviation and full word. Although some individual participants in the study showed a personal preference for one or the other, the interpreters as a group did not share a common trend. The inconsistencies concerning the choice between abbreviation and full word in note-taking can be subject to several factors, such as the language pair and the type of participants, but the empirical evidence to date is too thin to draw conclusive results.

Another choice that interpreters need to make during note-taking is the language of notes. Traditionally, the categories used to discuss this choice have been source and target languages. In the prescriptive stream of literature, while some publications have given a preference to the source language (e.g., Alexieva, 1994; Kirchhoff, 1979), others have recommended the target language (e.g., Herbert, 1952; Jones, 1998; Rozan, 1956/2002). More recently, researchers have found that the choice of language is also affected by whether a language is the native or non-native language. Empirical studies have been carried out to answer the questions centred on the choice of language. In general, student interpreters have been found to take their notes predominantly in the source language, but professional interpreters appear to have very varied language preferences (Abuín González, 2012; Andres, 2002; Dai & Xu, 2007; Lim, 2006; Liu, 2010; Lung, 2003; Wang et al., 2010).

It is important to note that both directions of interpreting need to be examined to simultaneously account for the source/target language factor and the native/non-native language factor. This has been achieved in very few studies and inconsistent findings have been reported. Dam (2004b) studied four student interpreters, of which three were Danish native speakers and one was a Spanish native speaker. The notes were dominated by the native language and she explained that it was because the students chose their better mastered language. Szabó (2006) studied eight professional interpreters with a language pair of Hungarian (native language) and English (non-native language). She found a contradicting preference for the non-native language, English, and pinpointed the morphological complexity and economy of writing as impacting factors. Wang et al. (2010) studied 12 student interpreters with a language combination of Chinese (native language) and English (non-native language). They found a source language dominance regardless of the direction of interpreting, and inferred that this resulted from the participants' inadequate interpreting competence.

This study analysed the notes taken by 22 professional interpreters while they performed CI between Chinese and English (see Chapter 5). The interpreters used significantly more English (the non-native language) than Chinese (the native language) in both directions of interpreting. A post-experiment questionnaire was used to capture the interpreters' personal account of their note-taking preferences (see Appendix D). Results of the questionnaire show that the interpreters prefer English because it can be written down faster, using phonetic spelling and even misspelling, whereas the written form of Chinese is not linked to its pronunciation in such a straightforward way. The study also found that the preference for English was stronger when English was the source language and that this preference was weakened when Chinese became the source language. Taken together, data in this study suggest that during note-taking, interpreters are subject to the influence from both the characteristics of the specific language pair and the source/target language status.

What is worth noting is that the participants in this study were recruited based on stringent criteria (see Chapter 1) and although they were native Chinese speakers, they were based in an English-speaking country (Australia). Therefore, the participants were experienced professionals with a strong non-native language, similar to the participants in Szabó's study. Their language competence and experience should also have played a role in how they chose the note-taking language.

The case of language choice in note-taking is far from closed because the empirical data is still limited and many of the previous findings have not been tested statistically. Even so, combining what has been found in literature and in this study, some tentative conclusions can be drawn. The choice of language in note-taking seems to be a function of the combined influence from a series of factors, including (1) the nature of the languages themselves (e.g., morphological complexity and economy of writing); (2) the native/non-native language status; (3) the source/target language status; and (4) interpreter characteristics (e.g., working experience and language competence). More specifically, when one of the languages is morphologically simpler and/or easier to write, this language would be the preferred choice regardless of the native/non-native language status, especially when the interpreter has a strong non-native language. When two languages do not differ too much in morphological complexity and economy of writing, the native/non-native language status plays a leading role in determining the language choice: interpreters prefer the native language. When the interpreting competence is inadequate (e.g., student interpreters at a beginning stage), the language choice is mainly subject to the source/target language status, with source language being the dominant language in note-taking.

### 7.1.2 Note-taking and cognitive load in CI

Literature suggests that the main concerns in note-taking are the time needed for writing, the effort of the hand and the cognitive costs (Alexieva, 1994; Allioni, 1989; Gile, 2009, p. 178). In this study, the temporal, physical and cognitive demands of varied note-taking choices were inspected via pen recording (see Chapter 5). The data suggest that interpreters' note-taking choices in Phase I of CI mainly correlate with the cognitive demand, rather than the temporal or physical demands.

With regard to the choice of form, the cognitive demand of language notes was found to be lower than that of symbol notes, and interpreters prefer language to symbol even though the physical and temporal demands of language notes were higher. Interpreters do not have a preference between abbreviations and full words, which are associated with similar levels of cognitive load, despite the fact that the physical and temporal demands of abbreviations are lower.

In terms of the choice of language, the cognitive demand of English notes was found to be lower than that of Chinese notes when interpreting from Chinese to English. In this case, interpreters prefer English to Chinese despite the facts that notes in the two languages are similar in their physical and temporal demands and that English is the non-source and non-native language. When interpreting from English to Chinese, no significant difference was found between the cognitive demands of the two languages. In this case, English seems to have become the preferred choice because of its source-language status.

In Phase II of CI, the cognitive demands of note-taking choices were inspected via eye tracking (see Chapter 6). The data suggest that language notes are more cognitively demanding to read than symbol notes. Notes in these two forms differ not only in early and fast cognitive processing such as recognising and identifying the note and its meaning, but also in slower operations such as linking the note to the context and producing a target speech around it. Moreover, eye movement measures including average fixation duration, number of revisits, and skip rate suggest that reading language notes and reading symbol notes can almost be treated as different tasks: the differences between them are greater than those found between different reading types (e.g., reading for comprehension and reading for translation).

The study also uncovered that full words are processed for longer times than abbreviations, not because they are more difficult to recognise or comprehend, but mainly because they are longer. In other words, the difference in visual attention received by abbreviations and full words is a result of the length factor. However, considering that time is a valuable resource in interpreting and that full words do cost more time to process during note-reading, abbreviations seem to be a more optimal choice because they attract less visual attention and take less time to process.

In addition, data in this study show that note-reading in Chinese is less cognitively demanding than note-reading in English when the direction of interpreting is from English to Chinese. In this case, two factors could have played important roles. First, Chinese is the native language of the interpreters, hence reading Chinese notes (native language processing) is likely to be less demanding than reading English notes (foreign language processing). Second, when interpreters read Chinese notes and produce a target speech in Chinese, reading and speech production are in the same language. This within-language processing is likely to be less cognitively demanding than between-language processing, i.e., reading one language and speaking in another language. This helps explain why the difference in cognitive demands between Chinese and English notes was not significant in the Chinese to English direction. When the direction of interpreting was reversed, the cognitive cost of reading Chinese notes was raised by between-language processing (from Chinese notes to English speech).

### 7.1.3 Note-taking and interpreting performance

Previous studies on the relationship between note-taking and interpreting performance have reached inconsistent conclusions (see Chapter 2). Some claim to have detected evidence pointing to a potential link between certain note-taking features and the quality of interpreting (Cardoen, 2013; Dam, 2007; Dam, Engberg, & Schjoldager, 2005; Her, 2001) while others fail to detect any such relationship (Dai & Xu, 2007; Wang et al., 2010).

This study inspected the relations between interpreting performance and a series of note-taking related factors, including the quantity of notes, the distribution of notes over form and language, and the time spent on multi-tasking (listening/analysing and note-writing at the same time) (see Chapter 4 and Chapter 5). In both directions of interpreting, a higher percentage of time spent on multi-tasking correlated with a better interpreting performance, whereas a higher percentage of English notes correlated with a worse interpreting performance. In addition, some correlations that were present in one direction of interpreting were absent in the other. In English to Chinese interpreting, the performance quality was higher when interpreters used more symbol notes and less language notes. In Chinese to English interpreting, the interpreting performance was better when the quantity of notes was higher. It has been made evident in this study that although some findings on the relation between note-taking and interpreting performance are consistent in both directions of interpreting, there are also critical differences between the two directions.

## 7.2 Note-taking, cognitive load and interpreting performance: the interplay and implications

The interplay between note-taking, cognitive load and interpreting performance are summarised in Table 7.1. Taken together, the data show that interpreters take proactive actions to reduce cognitive load during Phase I of CI, but the choices they make sometimes leave the extra burden to Phase II and are at the expense of interpreting quality. This means that the choices made by interpreters with the intention to ease the cognitive system are not always optimal when all factors are taken into account.

Table 7.1 The interplay between note-taking, cognitive load and interpreting performance

| | **Choice of form** | | **Choice of language** |
|---|---|---|---|
| | Language vs. symbol | Abbreviation vs. full word | Chinese vs. English |
| Physical and temporal demands | Language > Symbol | Abbreviation < Full word | Chinese ≈ English |
| Cognitive load in Phase I | Language < Symbol | Abbreviation ≈ Full word | Chinese > English[1] |
| Preferred note-taking choice | Language | No preference | English |
| Cognitive load in Phase II | Language > Symbol | Abbreviation < Full word | Chinese < English[2] |
| Relation with performance | Less language, more symbol, better performance[3] | No relation | Less English, better performance[4] |

Notes:

1. This result is only significant in the C-E task;
2. This result is only significant in the E-C task;
3. This result is only significant in the E-C task;
4. No significant correlation is found between the choice of Chinese notes and the interpreting performance.

When choosing the form of note-taking, interpreters would benefit more from using more symbols considering that they can save effort both in note-writing and note-reading, and that they have a positive contribution to interpreting performance. Interpreters could also use more abbreviations, especially considering that abbreviations save time during note-writing and that they are not more difficult to process than full words in note-reading.

In interpreter education, more emphasis should be put on teaching the students how to effectively use symbols and abbreviations. The training should have two outcomes: a decreased level of cognitive cost and an increased amount of usage. It is not enough to merely point to the highly individualised nature of symbols and abbreviations in note-taking and leave it to the students to develop a system of their own. The trial and error process is usually overly painstaking and students could potentially set out on the wrong track, even leading to under-developed note-taking skills towards the end of training. It is therefore advisable to provide a seminal set of symbols and abbreviations rich enough for the students to start with, together with rules on how to develop new symbols and abbreviations based on the given sample. The students can then spend more time on practising using the symbols and abbreviations instead of (re)inventing them.

When choosing the language of note-taking, many factors are relevant, such as the morphological complexity, economy of writing, native/non-native language status, source/target language status, and interpreting competence. There is no absolute correct answer when it comes to which language should be used. Specific strategies should be prepared to cater to different language pairs and stages of competence development.

What has also been made evident in this study is that there are some differences between the two directions of interpreting. Teachers, students and professional interpreters should mind this factor in their pedagogical planning, practices, and continued professional development. Diversified approaches should be taken to cope with the specific challenges brought by different interpreting directions.

## 7.3 Strengths and limitations of the study

Overall, the research has three main strengths. First, the research enhances the theoretical framework of studying the cognitive aspects of interpreting by presenting an in-depth discussion on the construct of cognitive load in interpreting and its measurement. Building on both translation and interpreting studies and the research on mental workload and Cognitive Load Theory, this study gives an operationalised definition to cognitive load in interpreting. A detailed discussion on the relevant variables and their interactions was presented. This discussion can not only help identify which factors potentially impact cognitive load in interpreting, but also provide a guideline in terms of which variables to control for to create better working conditions for interpreters and to obtain more valid experimental results in research. This study also elaborated on the measurement of cognitive load, especially how to measure the cognitive load in CI and note-taking, providing a good reference for future work to be carried out on the topic.

Second, this study makes a methodological contribution to language processing research (especially process-oriented translation and interpreting research) by triangulating the methods of pen recording, eye tracking, and voice recording. Through a careful design, diverse sources of data during the entire interpreting process were recorded in a synchronised way. This novel approach is likely to be beneficial to future investigations on the cognitive aspects of translation and interpreting, and more generally to language processing research.

Third, this research has contributed ample empirical data to reveal traces of processing efforts in note-taking and CI. It proposes a unique array of indicators for the measurement of the physical, temporal and cognitive demands of note-writing and note-reading, indicators that can be useful in future studies on related topics.

Despite the strengths outlined above, the research has several limitations. First, although a total of 26 professional interpreters participated in the study, valid data was collected from only 22 participants during pen recording (15% data loss) and 18 participants during eye tracking (31% data loss[1]). Purposive sampling was carried out through the professional network of the researcher, giving rise to exclusion bias and placing limits on the generalisation of the findings.

Second, this research only examined the correlations between the note-taking choices and a series of indicators of cognitive load. The current design does not allow any causal relations to be tested between the variables. Even though the findings are helpful for generating hypotheses for future studies, they cannot be taken as evidence of how cognitive load changes because of note-taking or the other way around.

[1] This kind of data loss is consistent with eye tracking studies of this nature (see, for example, O'Brien (2009) and Section 6.5.1).

Third, this study only involved one language pair (Chinese and English) and one type of interpreter (professional interpreters). It remains to be tested whether the findings can be generalised to other language pairs and to other types of participants.

Fourth, the eye tracker used in this study has a relatively low sampling rate (60 Hz) and is not compatible with the Eye and Pen software that is used to collect pen recording data. The synchronisation of the data requires laborious manual work. Although this particular eye tracker has been selected for a number of reasons (e.g., allowing free head and hand movement, high ecological validity, availability, see Chapter 1), the study would have benefited from an eye tracker with a higher sampling rate and compatibility with the Eye and Pen software.

## 7.4 The way forward

The findings and limitations of this research together have pointed to some interesting directions for future research. To begin with, it would be interesting to see how replicable this study is. A promising approach is to test whether the same results can be obtained using a different language combination or a different type of participant. A further step can be taken to see how the results change when tested with a cluster of participant groups with varying degrees of experience.

Another appealing topic is the causal relations between the variables investigated in this research. The observed correlations in this study form a good starting point for hypotheses to be formed. On the one hand, how does a change in note-taking choice affect the level of cognitive load? On the other hand, how will interpreters' note-taking choices change as a result of increased or decreased cognitive load?

A third issue that is worth investigating is other feasible cognitive load measures. Within the scope of a PhD research project, this study has tested the usability of a series of cognitive load measures. However, there are many other measures that are potentially useful for translation and interpreting studies. Explorations on whether certain measures are effective and how to apply the measures can be a productive line of research.

A fourth valuable avenue of research involves testing other apparatus and designs. Are there any other types of pen recording and eye tracking apparatus that would be useful in this kind of research? Can a different design increase the quality of data, reduce data loss, and cut down manual work in data analysis? Answers to these questions can help consolidate the foundations for experiments to be carried out on similar topics in the future.

Research on the cognitive aspects of interpreting is experiencing an exciting start. Hopefully more interested researchers will join the effort because this branch of interdisciplinary research can not only benefit the field of interpreting but also help enhance our understanding of language processing and the human mind.

## 7.5 References


Abuín González, M. (2012). The language of consecutive interpreters' notes: Differences across levels of expertise. *Interpreting*, *14*(1), 55–72. doi:10.1075/intp.14.1.03abu

Alexieva, B. (1994). On teaching note-taking in consecutive interpreting. In C. Dollerup & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 199–206). Amsterdam: John Benjamins.

Allioni, S. (1989). Towards a grammar of consecutive interpretation. In L. Gran & J. M. Dodds (Eds.), *The theoretical and practical aspects of teaching conference interpretation* (pp. 191–197). Udine: Campanotto.

Andres, D. (2002). *Konsekutivdolmetschen und notation [Consecutive interpreting and note-taking]*. Frankfurt: Peter Lang.

Cardoen, H. (2013). The Effect of Note-taking on Target-text Fluency. In G. González Núñez, Y. Khaled, & T. Voinova (Eds.), *Emerging Research in Translation Studies: Selected Papers of the CETRA Research Summer School 2012* (pp. 1-22). Leuven: CETRA.

Dai, W., & Xu, H. (2007). 汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例 [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research*, *39*(2), 136－144.

Dam, H. V. (2004a). Interpreters' notes: On the choice of form and language. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in Translation Studies* (pp. 251–261). Amsterdam: John Benjamins.

Dam, H. V. (2004b). Interpreters' notes: On the choice of language. *Interpreting*, *6*(1), 3–17. doi:10.1075/intp.6.1.03dam

Dam, H. V. (2007). What makes interpreters' notes efficient?: Features of (non-)efficiency in interpreter's notes for consecutive. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and directions in Translation Studies: Selected contributions from the EST Congress, Lisbon 2004* (pp. 183–197). Amsterdam: John Benjamins.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H. V. Dam, J. Engberg, & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (Vol. 7, pp. 227–254). Berlin: Mouton de Gruyter.

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

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation, 6*, 53-77.

Herbert, J. (1952). *Manuel de l'interprète: comment on devient interprète de conférences [Interpreter's manual: How to become a conference interpreter]*. Geneva: Georg.

Jones, R. (1998). *Conference interpreting explained*. Manchester: St. Jerome.

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang [The language of note-taking as a tool for the conference interpreter in consecutive interpreting]. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und sprachpraxis [Language theory and language practice]* (pp. 121-133). Tübingen: Gunter Narr.

Lim, H.-O. (2006). A post-mortem of note-taking. *Forum, 4*(2), 89–112.

Liu, J. (2010). 英语专业本科学生汉英交传笔记特征——一项基于学生交传笔记的实证研究 [Note-taking characteristics of English majored undergraduates in Chinese-English consecutive interpreting: an empirical study based on students' consecutive interpreting notes]. *Foreign Language World, 2*, 47–53.

Lung, R. (2003). Taking "notes" seriously in the interpretation classroom. In Á. Collados Aís, M. M. Fernández Sánchez, & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación [The evaluation of quality in interpretation: Research]* (pp. 199–205). Granada: Comares.

O'Brien, S. (2009). Eye tracking in translation process research: Methodological challenges and solution. In I. M. Mees, F. Alves, & S. Göpferich (Eds.), *Methodology, technology and innovation in translation process research: A tribute to Arnt Lykke Jakobsen* (pp. 251–266). Frederiksberg: Samfundslitteratur.

Rozan, J.-F. o. (1956/2002). *Note-taking in consecutive interpreting* (A. Gillies, Trans.). Cracow: Tertium Society for the Promotion of Language Studies.

Wang, W., Zhou, D., & Wang, L. (2010). 口译笔记特征与口译产出质量实证研究 [An empirical study of note-taking characteristics and output quality in interpreting]. *Foreign Language World*, *4*, 9–18.

Appendix A & B of this thesis have been removed as they may contain sensitive/confidential content

## Appendix C: Demographic questionnaire

*** 1. Participant ID**

*** 2. Professional background**

What is(are) your full-time job(s)

What is(are) your part-time job(s) (if any)

Which city do you most frequently work in as a consecutive interpreter? (Please name only one city)

How many NAATI working language combinations do you have? Please specify. (e.g. Two, Mandarin/English, and Cantonese/English)

How many years have you been practicing as a professional consecutive interpreter? (e.g. about 4 years)

Please give an estimate of the number of occasions you've provided consecutive interpreting service in the past 12 months. (e.g. about 12 occasions)

*** 3. What is the highest level of education you have COMPLETED?**

- ○ Bachelor's degree
- ○ Master's degree
- ○ Doctoral degree

Other (please specify)

*** 4. What kind of interpreting training and education have you received? (You may indicate more than one)**

- ☐ Intensive interpreting training course
- ☐ Interpreting diploma
- ☐ Postgraduate-level interpreting degree

Other (please specify)

*** 5. What time of the day do you function the best?**

**Appendix D: Post-experiment questionnaire**

Participant ID:

1. On a scale from 1 (extremely bad) to 7 (extremely well), how well rested are you before coming to the experiment? _____

| Extremely bad | | | | | | Extremely well |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

2. On a scale from 1 (never heard of) to 7 (extremely familiar), please rate how familiar you are with the topic of the source speeches. By familiarity, we mean how often you have come across the topic, as well as how well you know the topic and what it is about.
Topic of the Chinese speech: _____
Topic of the English speech: _____

| | Never heard of | | | | | | Extremely familiar |
|---|---|---|---|---|---|---|---|
| Topic of the Chinese speech: How to purchase property in Australia | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Topic of the English speech: How to register a business in Australia | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

3. On a scale from 1 (strongly disagree) to 7 (strongly agree), do you agree that the digital tablet and pen are sufficiently similar to real pen and paper, and therefore did NOT affect your note-taking behaviour? _____

| Strongly disagree | | | | | | Strongly agree |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

4. Have you received specialised training on note-taking? If so, do you think it has affected your note-taking style (e.g. use of symbol, choice of language)?

5. What is your preference on the form of note-taking? That is to say, do you usually use more language or symbol? Do you usually use more abbreviations or full words? Why?

6. What is your preference on the language of note-taking? That is to say, do you usually use more source or target language? Do you usually use more Chinese or English? Why?

7. Is the eye tracker (the glasses) comfortable? Do you think wearing the eye tracker has affected the main task of interpreting?

8. Do you think the instructions of the experiment are clear enough?

9. Do you have any comments on the interpreting tasks (e.g. topic, length, segmentation, difficulty, quality of audio)?

10. Do you have any other comments about today's experiment?

# Thank you very much!

## Appendix E: Scripts for the two interpreting tasks

Chinese to English task

Topic: 如何在澳大利亚买房

Segment 1

许多想在海外投资的人如今将澳大利亚视为排名第一的投资地。更好的气候，更清新的空气，买得起的房子以及大量的投资机遇带来的诱惑让人血脉偾张。

如果你已经决定要在澳洲买房，那么恭喜你。买房可以为你和你的家人提供未来的资产保障并创造财务上的自由。今天我将针对在澳洲买房谈一谈我的几点建议。

第一点建议，在买房前先要解决好法律上的问题。如果不是澳洲国籍，或者永久居民，你很可能要从澳洲的外国投资审查委员会获得批准才能购房。如果不想申请批准，不妨试试购买新建的房产。买新房不太需要经过审查，但下手要快，因为新房非常抢手。

Segment 2

第二点建议，找出澳洲哪里的房产最受欢迎。大城市比如悉尼、珀斯、布里斯班和墨尔本，这些地方的房价升幅最大。这些城市还可以提供最佳的工作机遇和较高的租房需求。除此之外，沿海地区由于旅游业的蓬勃发展，也可以提供获得收入的良好机会。这也是很多中国移民选择在黄金海岸和凯恩斯这样的热门旅游地做起生意的原因。如果你想让资金更细水长流一些，就可以看看更靠近郊区的地方。但要记住，郊区在澳大利亚基本就等同于偏远的地方，因此在着手购买前一定要做好研究。

第三点建议，尽可能多看一些房产。通过在多个地区看多套房产，你可以对各个地区及其房产的实际价值有更好的理解。你可以去网上注册几个房产网站。当你感兴趣的地区有房产可买的时候，这些网站就会提醒你。你可以重点关注那些已经在市场上销售超过六个星期的房产，或者是已经降价的房产。这表示购买这些房产的竞争压力比较低，并且你可能有讨价还价的余地。在你看过的房产中，你可以对其中的几个报价。可能你的报价中只有两三个会被接受，然后你就可以从中挑选你最满意的房产进行购买了。许多著名的房产投资人会为了购买一处房产而看一百多套房。

Segment 3

第四点建议，跟你的邻居聊聊天。多去走走，敲敲你邻居的门，跟他们说说话，从他们口中尽可能地获取当地的相关信息。有的时候这并不是要看邻居们具体说了什么，而是要看看你的邻居是谁。通过跟住在当地的人交流，你可以挖掘出很多当地街道和社区的信息，并了解到附近住的是什么样的人。

在澳洲买房尽管花费很高，但投资潜力是非常好的，尤其是在市区买房。举个例子，悉尼的房价在 2000 年一年增长了 50%。而东海岸在 1997 至 2003 年间的总增长更是达到了惊人的 112%。当然现在涨势已经没有那么凶猛了，但投资者不应因此丧失积极性。在西海岸的珀斯，房价仍然以每年 15%的速度在增长。如果打算将房子租出去，你的预计收入可能达到每年 3%到 6%。你还在等什么呢？

English to Chinese task

Topic: How to register a business in Australia

Segment 1

Today I'd like to share some information about registering a business in Australia. The number of Australians starting their own businesses is rocketing, and this entrepreneurial spirit looks set to continue in the near future. More than 520,000 new businesses have been registered during 2013, a rise of 8% on 2012 and a record high. The vast majority of those, at 95%, were small ventures launched by individuals. If you want to join them, today is your best opportunity to learn some useful information.

However, the first thing you need to know is that not everyone is entitled to register a business. Therefore, you must assess whether you are eligible to do so. These are some of the questions you need to ask yourself. Do you intend to make a profit from your activities? Is the activity of your business of a commercial nature? Are you committed to running a business? Your answer to these questions needs to be "yes" before you can register a business.

Segment 2

The second thing to know is the government organization that you'll deal with. The Australian Securities and Investments Commission, or ASIC, is an independent government body that regulates Australia's corporations, markets and financial services. It is the government body that you'll deal with the most when registering your business. The ASIC website is very convenient to use for various applications, and you will learn about many of its functions today.

To begin registering your business, you need to apply for an ABN. An ABN is short for Australian Business Number. It is a unique nine digit number, which helps government organizations identify your business. An ABN allows your trading partners to easily confirm your company's details for ordering and billing. This process can be done on your own, through a corporate service provider, or by an accountant on your behalf. If you decide to register on your own, you can do this either online or by lodging a form. If you register online, you usually receive your ABN immediately. However, some approvals may take up to twenty-eight days.

Once you have an ABN, you should do some business name research. Check the ASIC website to see if your desired name is available, because duplicate names will not be approved. For example, if you wish to register "Garden Services" as your business name, and this is already taken, you may need to try a slightly different name like "Jack's Garden Services".

Segment 3

The next step is to do some trademark homework to ensure that you are not infringing on anyone's intellectual property. Remember, the fact that a name is not registered doesn't mean it's not trademarked. It's also a good idea to check if your preferred domain name is available. You wouldn't want to have a web address too different from your business name.

Once you've done your research and decided on your desired name, it is time to apply. To apply for your business name, you will need to provide your ABN, your personal details such as date of birth and country, your email address, residential address, and business address. Business name registration is relatively simple. You can do it on your own through the ASIC website. Or if you don't like to deal with government organizations, you can have an authorized third party apply for you.

The turnaround time for business registration is about five working days. When your name is confirmed by the ASIC, you'll receive a letter of confirmation, which signals that you can register your domain, start creating your logo and business cards, and anything else you may need to do to get your new business up and running.

Hopefully you are now ready to take the next step. Good luck and I hope your new venture is successful.