The International Journal for Translation & Interpreting Research
trans-int.org

Note-taking in consecutive interpreting: New data from pen recording

Sijia Chen
Macquarie University, Australia
sijia.rachel.chen@gmail.com

DOI: 10.12807/ti.109201.2017.a02

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

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[i], 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*[ii], pen data during the note-taking process are recorded in great details. 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

[i] In this article, CI refers to long consecutive where systematic note-taking is used.
[ii] The website of the software is http://www.eyeandpen.net.

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.

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, 2004a; 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 (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, 1995/2009; 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, 2004a; Szabó, 2006; Wang et al., 2010).

Dam (2004a) 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, 1995/2009, 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 (1995/2009, 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?

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.

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).

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).

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 1).

Table 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 2).

Table 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

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.

3.5. Data and analysis

The data collected in this study are summarised in Table 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[iii] provide an emic perspective from the interpreters, enabling finer-grained analyses of the quantitative data, and help to explain the observed results.

Table 3: Data used for analysis

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

3.5.1. Categorisation of note units

Based on the interpreters' retrospection, all written notes were categorised according to their form and language (
Figure 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.

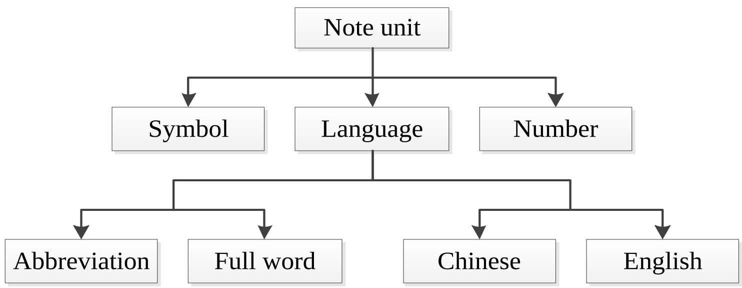

Figure 1: Categorisation of note units

The note categories and their definitions are specified in Table 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.

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

Table 4: Categories and definitions of note units (Adapted from Dam (2004a, p. 6) and Dam (2004b, 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[iv].	
Number	Independent from language and symbol, numbers are seen as a special category of notes.	

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 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).

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

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.

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 perfor-

-mance (e.g., Dam & Engberg, 2006; Gerver, 1969/2002; 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 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. 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.

4.1. Choice of form and language

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 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 5: Distribution over form: language vs. symbol

Participant	Total	Language	Symbol	Number
Task 1: Chinese to English				
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.

4.1.2. Choice between abbreviation and full word

There was a preference for abbreviation (34%) to full word (22%) averaged across participants and tasks (Table 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 6 Distribution over form: abbreviation vs. full word

Participant	Abbreviation	Full word
Task 1: Chinese to English		
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.

4.1.3. Choice of language

The distribution of notes over language (Table 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 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.

4.2. Relationship between note-taking and interpreting performance

As we can see in Table 8, neither the score of notes nor the quantity of notes alone could explain the variances in performance:

Table 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.

4.3. 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, 2004a; Liu, 2010; Lung, 2003; Wang et al., 2010) or professional interpreters (Andres, 2002; Dam, 2004b) 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, 2004b; 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 with what Dam (2004a) 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 (2004a)), 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.

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?

5.1. Pen data on the choice of form

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 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 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

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 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.

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 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 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

5.3. Interpreting the findings from a cognitive load perspective

No matter what choices interpreters make during note-taking, the basic question, as Gile (1995/2009, 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 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 11: How physical, temporal and cognitive demands affect interpreters' note-taking choices

	Form Language vs. symbol	 Abbreviation vs. full word	**Language** 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

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.

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*. 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 language. *Interpreting, 6*(1), 3-17. doi: 10.1075/intp.6.1.03dam

Dam, H. V. (2004b). 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. (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. *Meta, 36*(2-3), 431-439.

Gile, D. (1995/2009). *Basic concepts and models for interpreter and translator training. Revised edition*. Amsterdam: John Benjamins.

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

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*. Geneva: Georg.

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

Kirchhoff, H. (1979). Die notationssprache als hilfsmittel des konferenzdolmetschers im konsekutivvorgang. In W. Mair & E. Sallager (Eds.), *Sprachtheorie und Sprachpraxis* (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* (pp. 199-205). Granada: Comares.

Matyssek, H. (1989). *Handbuch der Notizentechnik für Dolmetscher*. 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*. 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.