

Can Computer-assisted Interpreting Tools Assist Interpreting?

Xinyu Wang, Caiwen Wang¹

Received: 29/01/2019

Accepted: 20/03/2019

Abstract

This experimental study explores how a potential computer-assisted interpreting tool affects consecutive interpreters' performance in accuracy and fluency and their cognitive process during interpreting. It consists of an experiment with ten participants and a follow-up interview with them. The results show that: (1) the tool has improved interpreters' interpreting accuracy but its impact on fluency is mixed; (2) the tool especially benefits those who have better language proficiency and have had interpreting experiences; and (3) the use of the tool has increased interpreters' cognitive load insomuch that novice interpreters who have relatively low language proficiency are adversely affected.

Key Words

computer-assisted interpreting tools, cognitive load, interpreting quality, consecutive interpreting, interpreting teaching



1. Introduction

This research is a direct response to Fantinuoli's (2018) call where he states:

¹ Xinyu Wang works as a freelance interpreter in China. Caiwen Wang is Senior Lecturer at the University of Westminster, UK. We thank the two anonymous reviewers for their extremely kind comments. All errors are ours. Email for correspondence: c.wang4@westminster.ac.uk

The emerging role of both process and setting-oriented [interpreting] technologies has started to be recognised by researchers and first studies on the subject have been published recently. Yet, the majority of studies is of a general or theoretical nature, while the number of empirical studies is still almost insignificant. However, in order to shed light onto the advantages and disadvantages of CAI tools, the way they are affecting the interpreting process and the tasks interpreters can perform better with their help and those which [they] cannot, research on new technologies needs to be performed not only on the basis of naturalistic methods (such as corpus analysis), but empirical experiments should be conducted also in stringently controlled experimental conditions (p. 170).

We propose that the integration of machine interpreting (MI) into human interpreters' workflow in interpreting mode may boost interpreting performance in accuracy by acting as a reference for interpreters, though at the same time it may adversely affect performance in fluency due to interpreters' cognitive overload.

2. Background

2.1. Interpreting and technology

Machines are assisting professional interpreters in many ways. Broadly speaking, CAI tools include those that are not designed specifically for interpreting, such as the internet, Word and Excel (Fantinuoli's 2018; Ortiz & Cavallo 2018). We will focus on those specifically designed for interpreters.

During the last fifteen years, many pieces of specialized CAI software have been developed to help interpreters. They mostly function as glossary building tools, which provide interpreters with better and quicker knowledge acquisition and glossary creation so as to speed up and optimise their preparation. Additionally, an increasing number of CAI tools can be consulted during the actual interpreting phase (though under time constraint) via the use of intelligent searching methods such as InterpretBank (Fantinuoli, 2016). Ortiz and Cavallo (2018) have carried out a detailed review of available CAI tools and categorised their use into three types: training, preparation and the actual interpreting phase, as well as according to whether or not they were developed specifically for interpreters. After having filtered out the tools that are 'no

longer available or have not been updated in 5 or more years' (p.18), they have produced a list of existing in-use CAI tools as in Table 1.

Table 1. List of updated CAI tools

Name	Category (main function)	Specific for interpreters	Training	Prep. (CDIP)	Prep. (Terminology)	simultaneous	consecutive	Remote interpreting	Replacement
Melissi (Black Box/VIE)	Training material	Y	X						
IRIS	Speech bank	N	X						
Moodle	Learning platform	N	X						
EU DG SCIC Speech Rep.	Speech bank	Y	X						
Interplex UE	Glossary management	Y			X	X			
InterpretBank	Glossary management	Y			X	X			
Interpreters' Help	Glossary management	Y			X	X			
Intragloss	Glossary management	Y			X	X			
Translated s.r.l	Glossary management	N			X	X			
BootCATT	Corpora building	Y		X					
SDL Multiter	Terminology extraction	N			X				

m Extract									
Simple Extract or	Terminology extraction	N			X				
Sketch Engine	Terminology extraction	N			X				
Termin us	Terminology extraction	Y			X				
TermSu ite	Terminology extraction	N			X				
Dragon NS	Speech recognition	N				X			
Evernot e	Note-taking	N					X		
Penulti mate	Note-taking	N					X		
Lecture Notes	Note-taking	N					X		
ZipDx	Audio conference	N						X	
WebEx	Audio and video conference	N						X	
Skype	Audio and video conference	N						X	
Capiche	Automatic translation of text	Y						X	X
Voxtec/ Phrasela tor	Speech-to-speech system	Y							X

(Ortiz & Cavallo 2018: 19)

For the purpose of our current experiment, we are interested in Speech Recognition tools and Speech-to-Speech systems shown in Table 1. As their

names suggest, these tools are for use during the interpreting phase. Speech Recognition, as in Capiche, is the process of converting human speech signals into a sequence of words by means of a computer programme (Jurafsky and Martin, 2009). Speech-to-speech systems like Voxtec are designed to replace human interpreters, so called machine interpreting (MI): ‘In these systems, a device captures human speech in a source language, searches for equivalents against pre-recorded sequences in a target language and reproduces such a sequence’ (Ortiz & Cavallo *ibid.*: 16). As shown in the above table, these tools are used in simultaneous interpreting. In our study, however, we will focus on the potential use of CAI tools for consecutive interpreting.

The pursuit of MI has been around for three decades (Arora *et al.*, 2013). The technology consists of three components: SR, MT and Speech Synthesizer. SR and MT are as explained in the above. Speech Synthesizer is the synthesized voice that reproduces the machine translation text and thus replaces the human interpreter. Evaluations (e.g. Sakamoto, *et al.*, 2013) show that MI has a relatively high accuracy rate when dealing with short conversations in the experimental environment. Machine Translation technology has experienced several developing stages since 1933 (Wu *et al.*, 2016). With the advancement of SR and MT technologies thanks to Artificial Intelligence, the performance of MI has constantly been improved.

Nonetheless, recent applications of MI in international conferences, such as Boao Forum 2018 and Translation Automation User Society Asia Conference 2018, have received rather negative feedback, as widely reported in media including Souhu (2018), Ijiwei.com (2018), cclycs.com (2018), etc. An infamous example is ‘road’ and ‘belt’ as in ‘the Road and Belt Initiative’ are translated as ‘road for transportation’ and ‘belt for the waist’ respectively. The unsatisfying performance of MI is mainly due to the flaws in SR and MT technology. One of the reasons that SR is difficult is that ‘natural speech can also change with differences in global or local rates of speech, pronunciations of words within and across speakers, and phonemes in different contexts’ (Gold *et al.* 2011: 60).

Ye (2017) has in particular shown the limitations of MT for the language pair English and Mandarin Chinese. Although the accuracy of MT between Chinese and English has shown a 60% increase since the integration of Neural MT into Google Translate, high quality MT is restricted within certain domains where a great number of standardized translation documents (produced by human translators) are available for the machine to learn.

We take the view that MI may not be able to replace human interpreters for reasons cited in Ortiz and Cavallo (2018):

Among the reasons [that] led [...] to claim that technology will not replace the interpreters in the future, there are: nuances, linguistic variation, non-verbal communication, accents, linguistic subtleties, emotion, understanding of the ‘between the lines’, flexibility of the human being’s adaptation, decision-taking, reliability, culture, metaphors, intonation, irony, ambiguities, unpredictability, capability of judgment (p.24).

At the same time, we take the view that if the outcome of MT and SR, which is part of MI, can be displayed on a screen for interpreters to refer to during the course of interpreting, the SR and MT technology in current speech-to-speech systems may benefit interpreters and interpreting quality. In other words, at the end of the MT process in a speech-to-speech system, the MT would not be solely relied on but merely referred to by the human interpreter and it would be the human interpreter rather than the voice of the speech synthesizer that would be heard. This means that the speech synthesizer would be disabled or removed if such a CAI reference tool were designed. Sun (2013) and Gong (2018) have considered the possibility of combining the first two steps of MI, namely, SR and MT, to create a CAI tool to assist interpreting. This is made even clearer in Feng (2018) where he distinguishes between computer-assisted interpreting and machine interpreting:

The rationale of computer-assisted interpreting is: the computer recognises a speech and translates it. The computer then displays the translation on a screen for interpreters to refer to. The difference between CAI and MI is thus whether the MT is used as a reference by human interpreters or re-produced by a synthesised voice. Such is the essential difference between CAI and MI. (translated from Mandarin Chinese by the authors)²

According to Pöchhacker, SR has ‘considerable potential for changing the way interpreting is practiced’ (2016:188). Fantinuoli, who is a productive researcher

² The source text for the quote is: 计算机辅助口译的原理是：计算机快速识别源语语音，然后将识别的内容进行机器翻译，最后将译文呈现在屏幕上，供译员参考。不难看出，最后的译文是显示在屏幕上供译员参考还是转换成语音由机器播出，这是机器口译与计算机辅助口译的根本区别之所在。

in CAI tools (e.g., 2016, 2017a, 2017b, 2017c, 2018), points to the use of SR for terminology building in that ‘Speech recognition [...] could represent the next step in the evolution of CAI tools. It could be used to automatically extract terminology in real-time from the interpreters’ database or to show name entities, numbers and the like on the interpreter’s monitor’ (2018:170-171).

In our research, we explore the use of SR combined with MT for interpreters to refer to during the interpreting phase.

2.2 Interpreters’ use of CAI tools and the Effort Model

While Computer-Assisted Translation (CAT) tools are widely known and used among translators, CAI tools seem to remain at their infant stage (Fantinuoli 2018), and ‘Unlike translators, interpreters have rarely benefited from language technologies and tools to make their work more efficient’ (Corpas Pastor 2018: 138). Paster and Fern (2016) have carried out a survey among professional interpreters to learn which technology tools they use and what their opinions are on these tools. The results show that among the 133 professionals surveyed, less than five used SR tools and no more than twenty-five consulted glossary or other forms of terminology database during interpreting. The survey also shows that interpreters all reported the struggle of time for processing when showsusing CAI tools during the course of their interpreting.

CAI tools at the moment seem like a double-edged sword. On the one hand,

[...] computers and new technologies offer potential for easing some of the transfer burdens related to interpreting tasks, in that they can help interpreters in their real-time work providing them with quick access to a broader range of information in electronic dictionaries, databases and glossaries (Tripepi Winteringham 2010: 90).

But on the other hand,

[...] is their practical use feasible and does rendition benefit? The main drawback of the use of these tools is that it is still considered, at least in the booth, to some extent as unnatural [...], presumably because it may be time-consuming and distracting in an activity that requires concentration and fast-paced decoding and delivery. The interpreter at work may not have the time or the cognitive ability to look up a word online or in his/her electronic dictionary, or detect and choose the correct translation [...] In addition, [...]

should the right word be found it may not be possible to incorporate it smoothly in speech (Tripepi Winteringham *ibid.*: 90-91).

Unlike Tripepi Winteringham (*ibid.*), who along with most scholars focuses on simultaneous interpreting, our current study chooses consecutive interpreting as our departure point to study the impact of CAI tools on interpreters.

As early as in 1995, Gile proposed two models, for simultaneous interpreting and consecutive interpreting respectively, to explain the cognitive process interpreters need to go through to complete an interpreting task. The models were revised in 2009. The theories centre on interpreters' cognitive effort and aim to explain recurring difficulties other than linguistic capacity in interpreting. These models are called Effort Models. As our current focus is on consecutive interpreting, we represent his consecutive interpreting Effort Model here.

The Effort Model explains the interpreting process in two phases (Gile 2009: 175-176). Phase One is the comprehension phase (or listening and note-taking phase) consisting of interpreters' efforts in listening and analysis, note-taking, short-term memory operations and coordination as follows:

Phase one: listening and note-taking

Interpreting = L + N + M + C (L: Listening and Analysis; N: Note-taking; M: Short-term memory operations; C: Coordination)

Phase Two on the other hand is devoted to the production of notes and consists of interpreters' efforts in remembering, note-reading, production and coordination as follows:

Phase two: target-speech production

Interpreting = Rem + Read + P + C (Rem: remembering; Read: Note-reading; P: Production)

According to Gile, interpreters tend to work near their processing capacity 'saturation' and beyond this, errors are likely to occur. Problem triggers are associated with 'increased processing capacity requirements which may exceed available capacity or cause attention management problems, or with vulnerability to a momentary lapse of attention of speech segments with

certain features' (ibid.: 171). '[T]he following conditions must be met at all times in order for consecutive interpreting to proceed smoothly' (ibid.: 176):

(1) $LR + NR + MR + CR \leq TA$

(2) $LR \leq LA$

(3) $NR \leq NA$

(4) $MR \leq MA$

(5) $CR \leq CA$

(R: required mental or cognitive operational processing capacity; A: available required mental or cognitive operational processing capacity; T: total)

It follows that when inequality (1) is not true, saturation may occur. When inequalities (2), (3), (4) or (5) are not true, failure may result in spite of the possibility of total available capacity being larger than total requirements (ibid.: 177).

For the obvious reason that CAI tools are a relatively new development, the effort on an interpreter's part to use a CAI tool or to refer to materials produced by a CAI tool is not included in Gile's Model. We contend that in the new era where interpreting embraces technologies, using a CAI tool consisting of SR and MT for consecutive interpreting will have consequences in Phase Two above, as the tool has implications for attention management and thereafter for interpreting quality. In terms of attention management, an interpreter will now need to split their attention and coordinate between their notes, their long-term memory and the MT for their interpreting planning before delivery. In terms of interpreting quality, it will have at least the following two implications:

(1) advantages: the tool will potentially aid memory, note-reading, and production, thus potentially help improve interpreting quality.

Or,

(2) disadvantages: the tool will be a potential 'problem trigger', because it adds the effort of reading and processing the MT, and this also demands more coordination, which will have implications for attention management among other aspects of problem triggers.

There are then two consequential possibilities: one is that interpreting quality is improved as an interpreter can take advantage of CAI tools, and the other is that interpreting quality is compromised as the interpreter finds it hard to coordinate between his/her cognitive efforts in Phase One and Two above and his/her efforts to use CAI tools, which subsequently may lead to errors. As interpreting is conducted under time constraint, the use of CAI tools may directly impact on accuracy and fluency, two important factors for assessing interpreting quality.

Given interpreters' reluctance or discomfort in using technologies and the extra cognitive load incurred, implications for interpreter training may need serious attention. As Fantinuoli points out, if CAI tools consistently show overall positive effects on the interpreting products of both interpreting students and professionals, "there is no reason why advantages and shortcomings of their use should not be properly addressed in the training of future interpreters" (2018:169).

The aims of our current study are to find out the impact of CAI tools on interpreting quality and interpreters' cognitive experience as well as the implications of such impact. For this purpose, we sought to conduct an experiment, as called for by Fantinuoli (2018; Also see Section 1).

We have identified in literature hitherto two relevant experimental studies of the impact of CAI tools on interpreting quality, one by Sun in 2013 and the other by Fantinuoli in 2017c.

Sun's experiment tested whether having reference materials during interpreting can help increase the accuracy of the interpreting product. Sun recruited three groups of research subjects with two participants in each. In her study, all participants were asked to interpret three identical speeches from Chinese into English under different conditions: Group 1 without any reference (control group), Group 2 with the MT of the whole ST acquired from Google Translate as reference and Group 3 with only some key terms in English or Chinese as reference respectively. The accuracy rate of interpreting for each group is respectively 82%, 89.5% and 94%. It can be seen that using references can indeed help interpreters improve their performance in accuracy.

However, Sun's seemingly positive results are not convincing enough for three reasons. Firstly, the number of the participants is relatively small, as each of the three experimental conditions was represented by only two interpreters and

thus the participants' different interpreting experience and level might greatly affect the result as Sun admitted (*ibid.*:24). Secondly, as Sun provided Group 2 with the Google MT of the original Chinese transcription of the ST, and Group 3 with the MT of every terminology, she did not pay much attention to the fact that due to technical limitations, SR software might not be able to transcribe the whole speech (as well as punctuation) correctly. Lastly, when assessing the accuracy of interpreting, neither tangible criteria nor the number of the marker(s) was elaborated, which raises the question of subjectivity in assessments. In our effort to enrich relevant data, our present study will also address these three methodological issues.

With regard to Fantinuoli's study (2017c), he conducted an empirical study to test the ability of InterpretBank, an SR-CAI tool, to retrieve terminologies from an existing glossary and to identify numbers, so as to test the feasibility of integrating SR in traditional glossary management CAI tools. The experiment involved four steps. First, the SR programme transferred the voice message into words. Next, the SR-CAI tool read and pre-processed the provided transcription, with which it queried the glossary to look for terms to match. Lastly, it displayed on the interpreter's monitor the results of identified terms as well as all the figures detected. This experiment showed a promising result for this SR-CAI tool: it successfully recognized 113 out of 119 terms (corresponding to 94.96%) and all eleven numerals (corresponding to 100%). Fantinuoli's experiment showed convincing positive results for the combination of SR and CAI tools, based on a relatively large amount of research data. It demonstrates the possibility of the SR-CAI tool to enhance interpreters' performance. However, this possibility cannot be confirmed until solid evidence is available. As Fantinuoli (2017c:33) points out at the end of his study:

further investigation would be necessary to evaluate its impact on the interpreting process and product. For example, it has to reveal whether the interpreter may experience a visual (and cognitive) overload when working with SR-CAI tools or if their use may lead to the expected quality increase in the interpretation of specialized texts.

Our current study was designed to assess whether or not interpreters may experience a visual and cognitive overload and whether or not their interpreting will be affected when they use SR-CAI tools during interpreting.

2.3 Summary and current research focus

Given the current status of relevant research, which calls for more empirical and experimental studies, our current experimental study was designed to test the impact of CAI tools on both interpreting quality and the cognitive process of interpreters. Specifically, we designed our experiment to test how referring to the whole MT text would affect the general quality of consecutive interpreting in terms of accuracy and fluency and how interpreters would feel about the use of technology in interpreting.

We propose two hypotheses: (1) When interpreting with a reference, participants would have a higher accuracy performance but a lower fluency performance than when they interpret without the reference; (2) When interpreting with a reference, participants would on average have a higher accuracy performance but lower fluency performance than those participants who interpret without the reference. We also sought to assess the cognitive efforts involved while interpreters used the reference during the course of their interpreting.

3. Methodology

3.1 Introduction

We chose to study the impact of CAI tools on accuracy and fluency because these two aspects are among the most important parameters for interpreting quality (e.g., Bühler 1986, Chiaro & Nocella 2004, Pöchacker & Zwischenberger, 2010, where ‘sense consistency with original message’ is used instead of ‘accuracy’), and also because we expected our tested CAI tool to have greater influence on them.

3.2 Method

The experiment includes a major quantitative research and a follow-up qualitative research. In the quantitative research, interpreting performance in accuracy and fluency is examined to assess the impact of the CAI tool on interpreting quality. In the qualitative research, i.e., the study of an in-depth interview with our participants, the characteristics of using the reference produced by the CAI tool during interpreting were analysed by following Hale

& Napier (2013), so as to understand interpreters' general opinions on using references while interpreting.

3.3 Participants

All participants are MSc degree students studying specialised translation and interpreting at a British university. They are all native Chinese speakers with good IELTS scores (which are one of the entry requirements for their taught course) and study interpreting between English and Chinese³. At the time of our study, they had finished their learning classes. Their background varies in terms of working experience and previous education experience, as in Table 2.

Table 2 Background information about research participants

Participant	Professional accreditation	Number of conferences interpreted for	Score of IELTS Overall (Listening/Reading/Writing/ Speaking)
CI01	No	0	7.5 (8/9/6.5/7)
CI02	No	0	8.0 (9/9/6.5/7.5)
CI03	No	0	7.5 (8.5/8/6.5/6.5)
CI04	No	0	7.5 (8/7.5/6/7.5)
CI05	No	0	7.5 (8.5/9/6.5/6.5)
CI06	Yes	5	8.0 (9/9/7/7)

³ To apply to study at a British university, an applicant must be able to provide recent evidence that their spoken and written command of the English language is adequate for the programme for which they have applied, if English is not their first language and they are not a national of a country deemed by either the UK Home Office or a university to which they have applied to be 'majority English speaking'. The IELTS (The International English Language Testing System) academic version is the preferred English language qualification for the British university in this research and the requirement is an overall grade of 7.5 out of the total 10, with a minimum of 6.5 out the total of 10 for each of the subtests, i.e., listening, speaking, reading and writing.

CI07	Yes, CATI 2 (China's national accreditation for interpreters)	10	8 (9/9/6.5/7.5)
CI08	No	0	7.5 (8.5/8.5/6.5/6.5)
CI09	No	2	8 (8.5/9/6.5/7.5)
CI10	No	0	7.5 (8.5/8.5/6/7)

This background information is one of the extraneous variables for our experiment and we believe it may explain some of the results we observed.⁴

We initially divided the participants randomly into two groups: Group 1 and Group 2, so that one group would interpret one speech with the MT reference while the other would interpret the same speech without the MT reference and the two groups would alternate when they interpreted the second speech. In our observation, however, one participant, CI06, did not use the given reference at all in her interpreting when her group was asked to use the reference. As a result, data of her interpreting was analysed with the group with no reference, which we will detail in Section 4.

To counterbalance the effect of extraneous variables and ensure the validity of experiment results, we used two speeches, namely E1 and E2 (See Section 3.4), of a similar level of difficulty. The experiment was then conducted in the way shown in Table 3:

Table 3 Assignment of research participants

	E1	E2
With MT reference	CI01, CI03, CI05, CI07, CI09	CI02, CI04, CI06, CI08, CI10
Without MT reference	CI02, CI04, CI06, CI08, CI10	CI01, CI03, CI05, CI07, CI09

⁴ Extraneous variables are independent variables that a researcher has no control over and that can interfere with the results of a study (Hale & Napier 2013:158).

3.4 Speeches for interpreting

The speeches chosen, E1 and E2, are about climate change, a topic that the participants had studied for interpreting in their taught course. Thus, they had a more or less similar level of familiarity with the content. The two speeches are two excerpts from a single speech by former President Barack Obama, who delivered it to leaders from the Pacific Island Conference and the International Union for the Conservation of Nature World Conservation Congress Hawaii in 2016 (YouTube, 2016). E1 is extracted from 3:05 to 5:35 while E2 is from 5:38 to 8:20. The two excerpts each last for approximately two and a half minutes and consist of 378 and 362 words, respectively. We believe the two parts are of a similar level of difficulty as they both contain 15 specialised terms, and six and five numerals respectively and the sentence structures and lengths are similar.

3.5 Experiment tools

Ideally, the CAI tool for this experiment would be an SR-CAI tool which can automatically display the whole MT. However, as no such a tool is available on the market, all automatic processes involved were done manually instead. As we discussed in Section 2.1, computer-assisted interpreting essentially involves the use of a machine translation text displayed to interpreters for reference. What we did in our experiment was therefore to display the machine translation texts on a computer screen and use ourselves as the actual tool for the implementation of the MT. We hope future technology will realise this tool.

We used Dragon Anywhere for speech recognition because it is one of the most frequently used and most accurate pieces of SR software (e.g., TechRadar.pro, 2017; Aiken, Park & Balan, 2010). We used Google Translate as our MT tool due to its high MT quality (Ye 2017) as well as its zero cost. Other consecutive interpretation necessities, including notepads, pens, and headsets, were provided. Word file on the computer was used to show the reference and each interpreting session was recorded by a recording pen.

3.6 Experiment implementation

3.6.1 SR software transcribing

Each speech was played on a computer and Dragon Anywhere (the SR software) on another device picked up the voice and transcribed the speech into texts. This process was repeated three times for both excerpts and the most accurate result was used for our experiment. The error rates of the SR are presented in Table 4.

Table 4 SR error rates

	Error rate - E1	Error rate - E2
1 st time	4.08%	6.62%
2 nd time	4.96%	4.42%
3 rd time	4.12%	6.94%

After comparison, we used the first SR transcription of E1 and the second SR transcription of E2 for our experiment.

3.6.2 Pilot experiment

We ran a pilot experiment first for our study.

Conducting pilot studies can improve the reliability of experiment design. A well-run pilot study with a few participants can ensure that useful data are generated in the actual data collection that follows (Mellinger & Hanson, 2017:195).

In our pilot experiment, only one participant, CI02, was invited. The experiment material was a one-minute (00:01-00:58) excerpt from another speech about Climate Change (YouTube, 2017). We followed all experiment procedures mentioned above to produce the reference. CI02 was asked to interpret the speech with the corresponding MT reference. After the pilot experiment, CI02 stated that there was disarray at the beginning because not

enough information about the experiment had been provided. As a result, we decided to give clear and consistent instructions in our actual experiment, to ensure that participants were clear about what they were supposed to do. Every time before we started our experiment, the instructions were read out (See Section 3.6.3).

CI02 also suggested that the corresponding reference should be displayed sentence by sentence or utterance by utterance instead of in big paragraphs (as shown in Figure 1). However, as the current status of the SR and MT software is such that they display transcripts and translation texts in long paragraphs, we adhered to our original plan, which is to present the references as they were automatically generated by the SR software and then echoed by the MT. As we show in Section 4, most of our participants voiced similar opinions about the format of the references in our interviews with them, which as we discuss in Section 5 may point to a direction for future CAI.

从缅甸州到俄亥俄州到加利福尼亚州的指定国家保护区，就在上周，由于这个房间里很多人的辛勤工作，包括审查，我创造了世界上最大的海军陆战队，使纪念碑大小增加了三倍，看起来不错，这个区域的面积是原来的两倍。德克萨斯州将受到保护，并允许我们保存和研究明天受气候变化威胁的脆弱生态系统，前往中途岛环礁自己看看7000种生活在四分之一的物种，这些物种在我们周围世界其他任何地方都找不到。包含了今生之间的界限，接下来的数百名勇敢的美国人那里献出自己的生命来捍卫世界的自由，所以这个神圣的地方值得这样对待，从现在开始，它将为后代保留，为他的成就感到骄傲但我们总是进入小细胞，没有一个国家可以做到这一点就必须共同解决最终的变化像许多人这样的国家显然感到特别紧迫关于像我国和中国这样的最大排放国有特殊责任，确保各国尽自己的一份力量通过肮脏的发展阶段，转变为清洁能源战略，这是巴黎协议的一项关键原则，我在这里向领导人提及如果没有岛屿国家的不懈努力和辛勤工作，我们就无法获得巴黎协议，他们为加快公私清洁能源创新做出了巨大改变，确保资源将用于那些需要帮助准备的国家。气候变化的影响，我们再也无法负担了这一点也是为什么今晚我可以宣布超过3000万美元的新朋友承诺和为可持续发展和安全饮用水加强基础设施等投资提供资金

Fig 2: reference for E2

3.6.3 Interpreting session

During our experiment, we read the following written instruction to each participant at the start of their interpreting session:

This experiment aims to find out whether Speech Recognition and Machine Translation tools can help boost interpreting performance. You will be interpreting two excerpts from a speech about Climate Change, one with reference, the other without reference. Both excerpts last for around two and a half minutes. The reference is the Google MT of the whole speech transcribed by a SR tool with terminologies highlighted in yellow and numerals in green. However, while all terminologies highlighted are trustworthy since they were extracted from a glossary, numerals dictated by the SR tool may be wrong. The reference will be shown on the screen right after the excerpt finishes and you have to start interpreting as soon as the reference appears. You are allowed to make notes during interpreting. Your interpreting products will be assessed by both accuracy and fluency.

The experiment started after the participants confirmed that they understood the instruction.

Additionally, a written brief, which reads ‘Former President Obama gives this speech to leaders from the Pacific Island Conference and the International Union for the Conservation of Nature World Conservation Congress in Hawaii in 2016’, together with the term ‘Papahānaumokuākea National Monument’ paired with its Chinese translation ‘帕帕哈瑙莫夸基亚国家海洋保护区’, was showed on a paper to each participant before the experiment. All participants were given a minute’s time to get ready.

At the completion of each speech, the MT reference appeared on the computer screen for the participants to use for interpreting. Each interpreting was recorded by a recording pen.

During the experiment, when asked to use the MT as the reference, none of the participants who used it chose to only use the texts by reading them from the screen. Instead, they were seen to use both their notes and the texts on the screen while they were interpreting.

3.6.4 In-depth interview

We conducted an interview with each participant immediately after their interpreting session. This was to ‘gain an understanding’ into the interpreting experience of all participants as well as ‘the meaning they make of that

experience' (Hale and Napier, 2013:96). We adopted a 'semi-structured' interview, because it could enable the researcher to strike a balance between having some level of control and having some flexibility at the same time (Hale and Napier, *ibid.*).

Our interview questions are as below:

1. Under which condition do you think you performed better, with or without the reference?
2. Did you find the reference increased or reduced your interpreting performance?
3. Did the reference distract you from interpreting?
4. Did you use any terminology highlighted in the reference during interpreting?
5. Did you use any numerals highlighted in the reference during interpreting?
6. Did you use any part that is not highlighted in the reference during interpreting?
7. Did you find the reference reliable?
8. How can we improve the reference?
9. Please feel free to add anything.

Of note is that in both our instructions for the interpreting session and our questions for the semi-structured interview, we prominently mentioned terminology. This is part of a larger research project and in our current paper, we will not analyse the data on terminology.

All interviews were recorded with a recording pen and then transcribed for analysis.

3.7 Interpreting assessment criteria

According to Hale and Napier, 'Any research, but particularly quantitative experimental research, needs to be reliable and valid' (2013:160). As such, how to ensure the reliability and validity of interpreting assessment for accuracy and

fluency was our major concern in the present study. To ensure reliability, ‘test-retest’ method (Burns, 1997, cited in Hale and Napier, 2013:161) was applied by involving more than one assessor (Ibid.: 162). The first and the second author assessed the participants separately and independently before their discussion to agree on their ratings. To ensure validity, assessment criteria for both accuracy and fluency were carefully written by referring to a series of established quality metrics for both interpreting and translation, including two sets of marking criteria for interpreting used in two London universities. Besides, the criteria for accuracy were closely tailored to the speech. Each participant’s interpreting performance was assessed with a 1-5 score band.

Data out of the assessment were then analysed and discussed.

4. Data Analysis

4.1 Introduction

Data generated from the experiment session were used to test our hypotheses. Data from the interview were used to explore the rationale of the experimental results, though we also took our observations into account.

Each participant’s interpreting performance under the two conditions --- with and without the MT reference --- was compared. The interpreting of each of the two speeches under the two conditions was also compared. Table 5 presents the overview of our data, where “r” stands for “interpreting with the reference”, “Y” stands for “yes”, “N” stands for “no”, and “E” stands for “equal”.

Table 5 Overview of interpreting performance

Participant	Excerpt	Accuracy	Higher accuracy with reference?	Fluency	Lower fluency with reference?
CI01	E1 (r)	3.0		3.5	
	E2	2.7	Y	4.0	Y
CI02	E1	3.7	Y	3.0	E

Can Computer-assisted Interpreting Tools Assist Interpreting?

	E2 (r)	3.7		3.0	
CI03	E1 (r)	3.7		3.8	
	E2	2.9	Y	3.5	N
CI04	E1	1.3		2.8	
	E2 (r)	2.4	Y	2.0	Y
CI05	E1 (r)	3.7		3.8	
	E2	3.7	E	3.8	E
CI06	E1	4.7		4.8	
	E2	5.0	NA	4.8	NA
CI07	E1 (r)	4.9		3.5	
	E2	3.3	Y	4.0	Y
CI08	E1	2.7		3.5	
	E2 (r)	1.0	N	4.3	N
CI09	E1 (r)	2.7		3.0	
	E2	2.7	E	3.0	E
CI10	E1	3.0	Y	3.0	N

4.2 Data analysis to test hypotheses

4.2.1 Data analysis for the first hypothesis

To test our first hypothesis, we compared each participant's performance, in terms of both accuracy and fluency, under the condition of interpreting with

and without the MT reference (excluding CI06 for the reason we mentioned in Section 3.3).

As can be seen in Table 5, out of the nine participants, five (CI01, CI03, CI04, CI07, CI10) have better accuracy performance when interpreting with the reference, three have no change in their performance (CI02, CI05, CI09) and only one (CI08) is affected reversely. This means approximately 56% of those interpreters who referred to the MT text benefit from it. With regard to fluency, three participants (CI03, CI08, CI10) have better performance, three (CI01, CI04, CI07) have worse performance and three do not have any change. This result is therefore not conclusive regarding the consequence of using the CAI tool.

It is therefore clear the first half of our first hypothesis is confirmed by our experiment whereas the second half is not.

4.2.2 Data analysis for the second hypothesis

To test our second hypothesis, we compared the participants' performance according to the speech they interpreted. For easy comparison, we re-organised the data in Table 5, and presented it in Table 6 and Table 7 below, respectively. We used the average score for comparison.

Table 6: Interpreting performance for E1

Interpreting with reference			Interpreting without reference		
Participants	accuracy	fluency	Participants	accuracy	fluency
CI01	3.0	3.5	CI02	3.7	3.0
CI03	3.7	3.8	CI04	1.3	2.8
CI05	3.7	3.8	CI06	4.7	4.8
CI07	4.9	3.5	CI08	2.7	3.5
CI09	2.7	3.0	CI10	3.0	3.0
Average	3.6	3.52	Average	3.08	3.42

Table 7: Interpreting performance for E2

Interpreting with reference			Interpreting without reference		
Participa nts	accuracy	fluency	Participants	accuracy	fluency
CI02	3.7	3.0	CI01	2.7	4.0
CI04	2.4	2.0	CI03	2.9	3.5
CI08	1.0	4.3	CI05	3.7	3.8
CI10	3.3	3.3	CI06	5.0	4.8
			CI07	3.3	4.0
			CI09	2.7	3.0
Average	2.60	3.15	Average	3.38	3.85

In interpreting E1, the average accuracy score when interpreting with the reference was 3.6, whereas that for interpreting without the reference is only 3.08. The fluency score was also higher when interpreting with the reference. This result supports our second hypothesis in terms of accuracy performance but is the opposite of our hypothesis in terms of fluency performance.

In interpreting E2, the picture is different: the results support our second hypothesis in terms of fluency but is the opposite of our hypothesis in terms of accuracy performance.

The difference between results from E2 and those from E1 for the scenarios of interpreting with and without the MT reference was puzzling to us at first. However, when we looked at the participants' background as in Table 2, it seems the differences in IELTS scores and professional interpreting experiences between the two groups who interpreted E2 respectively with and without the reference provide the explanation. For convenience, we extracted relevant information from Table 2 and put it in Table 8 below:

Table 8: comparison of participants' background

Participants interpreting E2 with reference				Participants interpreting E2 without reference			
	iELTS	listening	Interpreting experience		iELTS	listening	Interpreting experience
CI02	8	9	0	CI01	7.5	8	0
CI04	7.5	8	0	CI03	7.5	8.5	0
CI08	7.5	8.5	0	CI05	7.5	8.5	0
CI10	7.5	8.5	0	CI06	8	9	5 conferences
				CI07	8	9	10 conferences
				CI09	8	8.5	2 conferences
Average	7.63	8.5	0	Average	7.75	8.58	3

As can be seen, the group who interpreted E2 with the reference have a lower iELTS score and a lower score in their listening element (which is important for interpreting), and none of them has had any professional interpreting experience. We believe these two extraneous variables caused the differences in the participants' performance in that the group interpreting E2 with the MT reference were less proficient in English and less experienced in interpreting, and therefore were less able to handle the cognitive load needed for interpreting with the CAI tool. This kind of cognitive overload was reported by the participants themselves, as shown in Table 9 (Section 4.3). Referring back to the Effort Model (Gile 2009; Also see Section 2.2), it can be argued that the efforts demanded from the participants for reading the MT and coordinating this with their other efforts in Phase One and Phase Two of the cognitive

process for consecutive interpreting exceeded the participants' available capacity, leading to the consequence that this group of interpreters did not perform well, not in accuracy and nor in fluency.

By contrast, the group who interpreted E2 without the reference score better in both their iELTS and their listening element, and three of them are experienced interpreters. This means that these participants are better at handling the cognitive load needed for the interpreting task and therefore they performed relatively well even without the reference. As a matter of fact, this same group also performed better when interpreting E1 with the reference. In particular, CI06 and CI07, who both have very high scores in their iELTS and the listening element and are most experienced in interpreting, performed very well in their interpreting of E1 and E2, whether or not they used the reference, and that they perform better when interpreting with the reference.

On this basis, our tentative conclusion is that the CAI tool may especially benefit those who have high language proficiency levels and who are experienced interpreters.

4.2.3 Summary

Overall, our data analysis indicates that on the part of the interpreters, they performed better in accuracy when interpreting both E1 and E2 if references were provided, and that on the part of the speeches, interpreting E1 with references generated better performance than without, but interpreting E2 with the reference was not as good as without. While this tendency largely supports our initial hypothesis that references or interpreting technologies help improve interpreting accuracy, our study shows that whether or not using MT as a CAI tool boosts interpreting quality also has to do with users' language proficiency and professional interpreting experience in that those who are more proficient in languages and are more experienced in interpreting will be able to take advantage of the CAI tool. Additionally, our study does not show any conclusive tendency for how the CAI tool may affect interpreters' fluency in interpreting.

4.3 Interview data analysis

As we indicated in Section 3.6.4, our interviews with the participants were intended to find out their experience with and confidence in using interpreting

technologies. After we have transcribed the participants' answers to our interview questions 1, 2, 3, 7, 8 and 9, we analysed and categorised their answers. Table 9 below provides an overview of the data analysis results.

Table 9 overview of data generated from interview

The tool and interpreting performance	<ul style="list-style-type: none"> • Participants thinking they performed better with the tool: CI02, CI03, CI09 (3/9) • Participants thinking they performed better without the tool: CI01, CI04, CI05, CI07, CI08, CI10 (6/9)
The tool being a distraction or reference	<ul style="list-style-type: none"> • Participants thinking the tool is a distraction: CI01, CI03, CI04, CI05, CI06, CI07, CI08 & CI10 (8/10); they at the time say technologies have potentials if interpreters are trained to use them. • Participants thinking the tool is a useful reference: CI01, CI02, CI03, CI04, CI05, CI06, CI07, CI09 & CI10 (9/10). CI02, CI03 & CI09: the speech is hard, and thus the MT helps, esp. the highlighted terms; if speeches are hard, then MT is helpful; technology is good for interpreters. CI03 & CI07: MT provides support if one misses things in their own notes. CI03 & CI09: MT as reference provides psychological support.
Anything to improve in the tool	<ul style="list-style-type: none"> • CI01, CI03, CI04, CI05, CI06, CI07, CI08, CI09 & CI10 (9/10): presenting the MT in one paragraph is not good, as it was difficult to locate needed information. • CI01, CI04, CI05, CI06 & CI10 (5/10): although MT is helpful, it was hard to locate the info in the MT. • CI03, CI04, CI05, CI08 & CI10 (5/10): the tool should chunk the MT text by sentence or utterance, so that interpreters can go for help if unsure about their own notes

As seen, three out of nine (approximately 67%) participants do not think they performed better when using the reference even though they did perform better (See Table 5), and eight out of ten (80%) think the tool is a distraction. This means that most of the participants do not have confidence in using the tool. The reason provided for the tool being a distraction is that the use of CAI tool during the course of interpreting made it hard for them to manage attention.

On the other hand, those in the minority who believe they performed better with the reference think that the CAI tool is helpful especially if they have missed anything in their own notes, and that knowing the CAI tool is available gives them psychological support, especially if a speech turns out to be hard to interpret.

Nine out of ten (90%) participants think having the tool would be useful and these participants include almost all of those who think the tool is a distraction at the moment.

Additionally, nine out of ten (90%) participants are not happy with the format where the text is displayed in long paragraphs, and according to CI06, this is the reason that she decided not to use the reference at all. The participants reported this format makes it very hard for them to locate the information they needed for interpreting. We believe this is the main reason for the low or lower fluency scores of some participants who interpreted with the reference, because the effort to locate the needed text would cause hesitations and pauses. The participants suggested that the text should be chunked to display sentence by sentence or utterance by utterance, as this would make it easier for them to locate needed information for interpreting.

5. Discussion and conclusion

We conducted our experiment under stringently controlled conditions in our attempt to assess the impact of CAI tools on interpreting and interpreters.

Our study has shown the tendency where CAI tools can improve the accuracy element of interpreting quality and can especially benefit interpreters with high language proficiency levels as well as professional experiences. On the other hand, the effect of CAI tools on fluency is inconclusive from our current study. We believe however that should the displaying format of references be improved, the tool would be able to improve this situation. We have also found that interpreters generally see CAI tools as a distraction as their efforts need to be split to process what a CAI tool has provided.

On the basis of our experimental study, we make the following suggestions for the future of CAI tools:

(1) As speech recognition and machine translation combined can potentially help improve interpreting quality during the course of interpreting, it is

promising to further develop relevant CAI technologies for this purpose. In our experiment, we drew inspirations from speech-to-speech systems (which are designed to replace human interpreters, though unsuccessfully), but manually implemented the procedure of speech recognition and machine translation as there is no such a product on the market that is ready to use. It would be encouraging to see such a CAI tool in real. Besides, the format of machine translation texts needs improving. The texts transcribed by the SR software and then generated by Google Translate are displayed in big paragraphs. This format made it hard for our participants to locate the information they needed for interpreting. As suggested by our participants, texts displayed in chunks by sentence or utterance would be particularly helpful for them to locate the needed information. Also, the MT texts, as well as the transcripts by the SR software do not display all needed punctuation. It would be good if this could be improved in future tool developing. In our belief, improvement in displaying format and punctuation of SR software would potentially improve interpreting fluency and help more with accuracy.

(2) There is a need for interpreting students and practitioners to become familiar with various CAI tools. Our study echoes literature findings that interpreters are generally uncomfortable with CAI tools. Our preliminary research of university interpreting courses indicate that the current status of interpreting teaching or interpreter training largely remained the same as before CAI tools came into being. This picture is different from that of translation, where teaching and learning courses provide training on how to use CAT (computer-assisted translation) tools, and as a matter of fact translators may find it hard to survive if they are not able to use CAT tools in their translation career. We believe similar training courses would help interpreters to become comfortable with and confident in using CAI tools, which would further help boost interpreting quality.

(3) Our study has shown a relationship between computer-assisted interpreting quality and interpreters' linguistic proficiency along with their professional interpreting experiences in that the better interpreters' language skills are and the more experienced they are, the better they are able to handle the extra cognitive load incurred by CAI tools so that the tools can be used for their benefit. While such a relationship is not surprising as those having better language skills and more experiences are generally supposed to be better interpreters with or without the help of CAI tools, our finding suggests the

need for anyone who wants to pursue a career in interpreting to continuously improve their language skills and gain professional experiences.

Although at the moment CAI tools are largely immature for use in the interpreting phase and interpreters experience extra cognitive difficulties and feel unnatural when using them, we believe sooner or later CAI tools will turn out to have more advantages than disadvantages for interpreters to embrace and benefit from.

References

- Aiken, M., Park, M., Balan, S., 2010. A Prototype System For Machine Interpretation. Available at: <http://translationjournal.net/journal/53mi.htm>
- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231-235.
- Cclycs.com. 2018. 译见 || 博鳌论坛 AI 口译演砸了, 人工翻译扳回一局? 专家怎么看? Available at: <http://www.cclycs.com/z610515.html>
- Chiaro, D., & Nocella, G. (2004). Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the World Wide Web. *Meta*, 49(2), 278-293.
- Corpas Pastor, G. (2018). Tools for Interpreters: the Challenges that Lie Ahead. In *Current Trends in Translation Teaching and Learning E*, 5, 157-182.
- Fantinuoli, C., 2016. InterpretBank. Redefining computer-assisted interpreting tools. In *Proceedings of the 38th Conference Translating and the Computer*. 42-52. London, UK. AsLing
- Fantinuoli, C. 2017a. Computer-assisted interpreting: challenges and future perspectives. In: DURÁN, IS.; CORPAS, G. *Trends in e-tools and RESOURCES for translators and interpreters*. Leiden: Brill, pp. 153-174.
- Fantinuoli, C. 2017b. Computer-assisted preparation in conference interpreting. *Translation & Interpreting*, Vol. 9, No. 2, 24-37. Available at: <http://www.trans-int.org/index.php/transint/article/view/565/294>
- Fantinuoli, C. 2017c. Speech Recognition in the Interpreter Workstation. In *Proceedings of the 39th Conference Translating and the Computer*, Geneva, 25-34. Available at: <http://www.staff.uni-mainz.de/fantinuoli/download/publications/Speech%20Recognition%20in%20the%20Interpreter%20Workstation.pdf>
- Fantinuoli, C., 2018. Computer-assisted interpreting: challenges and future perspectives. In *Trends in E-Tools and Resources for Translators and Interpreters*, 153-174. Koninklijke Brill Nv, Leiden

- Feng, J. (冯建中). 2018. 译术||机器口译：口译教育面临的挑战与对策. In: 翻译教学与研究. Available: http://www.sohu.com/a/247079088_176673
- Gile, D. 1995. Basic concepts and models for interpreter and translator training. Amsterdam: John Benjamins
- Gile, D. 2009. Basic concepts and models for interpreter and translator training. Amsterdam: John Benjamins
- Gold, Ben, Morgan, Nelson, & Ellis, Dan. (2011) Speech and Audio Signal Processing: Processing and Perception of Speech and Music. Hoboken, NJ, USA: John Wiley & Sons
- Hale, S. & Napier, J. (2013). Research methods in interpreting: A practical resource. London: Bloomsbury Academic.
- Hanson, T.A. and Mellinger, C.D. (2016). Quantitative research methods in translation and interpreting studies. London: Routledge.
- Ijiwei.com. (2018). AI 口译官发展史 . Available at: <https://www.laoyaoba.com/newinfo?id=673780>
- Jurafsky, D. & Martin, J. H. 2009. Speech and Language Processing (2nd Edition), Prentice-Hall, Inc. Upper Saddle River, NJ, USA
- Ortiz, L. & Cavallo, P. (2018). Computer-Assisted Interpreting Tools (CAI) and options for automation with Automatic Speech Recognition. *TradTerm*, Vol. 32, 9-31.
- Paster, G. & Fern, L. (2016). A Survey of Interpreters' Needs and Practices Related to Language Technology. Available at: https://www.researchgate.net/publication/303685153_A_survey_of_interpreters'_needs_and_practices_related_to_language_technology
- Pöchhacker, F. (2016). Introducing interpreting studies. London: Routledge.
- Pöchhacker, F. & Zwischenberger, C. (2010). Survey on quality and role: conference interpreters' expectations and self-perceptions. Available at: <https://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1>
- Souhu. (2018). AI 口译演砸了，翻译界专家怎么看? Available at: http://www.sohu.com/a/228313720_199210
- Sun, Y. (2013). The Supporting Role of Speech Recognition and Machine Translation in Interpretation. MA dissertation. Shandong Normal University.
- Techradar.pro. (2017). The best voice recognition software of 2017. Available at: <https://www.techradar.com/uk/news/the-best-voice-recognition-software-of-2017>
- Tripepi Winteringham S. (2010). The usefulness of ICTs in interpreting practice. In *The Interpreters' Newsletter*, Vol. 15, 87-99. Available at: <https://www.openstarts.units.it/bitstream/10077/4751/1/TripepiWinteringhamIN15.pdf>

Wu, Y., Schuster, M., Chen, Z., Le, Q. & Norouzi, M. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. Available at: <https://arxiv.org/pdf/1609.08144.pdf>

Ye, Z. (2017). 又想起了斯坦纳——写在新版谷歌翻译刀诞生之际. *中国翻译*, 38(1), 89-91

YouTube. (2016). Obama In Hawaii- Full Speech On Climate Change. Available at: <https://www.youtube.com/watch?v=2rRn01i8DpE&t=500s>

YouTube. (2017). Causes and Effects of Climate Change | National Geographic. https://www.youtube.com/watch?v=G4H1N_yXBIA

(The last access to all online references is December 2018)