

# Interpreting and technology

Edited by

Claudio Fantinuoli

Draft  
of October 22, 2018, 17:43

Translation and Multilingual Natural  
Language Processing 42



## Translation and Multilingual Natural Language Processing

Editors: Oliver Czulo (Universität Leipzig), Silvia Hansen-Schirra (Johannes Gutenberg-Universität Mainz), Reinhard Rapp (Johannes Gutenberg-Universität Mainz)

In this series:

1. Fantinuoli, Claudio & Federico Zanettin (eds.). New directions in corpus-based translation studies.
2. Hansen-Schirra, Silvia & Sambor Grucza (eds.). Eyetracking and Applied Linguistics.
3. Neumann, Stella, Oliver Čulo & Silvia Hansen-Schirra (eds.). Annotation, exploitation and evaluation of parallel corpora: TC3 I.
4. Czulo, Oliver & Silvia Hansen-Schirra (eds.). Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation: TC3 II.
5. Rehm, Georg, Felix Sasaki, Daniel Stein & Andreas Witt (eds.). Language technologies for a multilingual Europe: TC3 III.
6. Menzel, Katrin, Ekaterina Lapshinova-Koltunski & Kerstin Anna Kunz (eds.). New perspectives on cohesion and coherence: Implications for translation.
7. Hansen-Schirra, Silvia, Oliver Czulo & Sascha Hofmann (eds.). Empirical modelling of translation and interpreting.
8. Svoboda, Tomáš, Łucja Biel & Krzysztof Łoboda (eds.). Quality aspects in institutional translation.

# Interpreting and technology

Edited by

Claudio Fantinuoli

Claudio Fantinuoli (ed.). 2018. *Interpreting and technology* (Translation and Multilingual Natural Language Processing 42). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/209>

© 2018, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2364-8899

no DOI

Source code available from [www.github.com/langsci/209](http://www.github.com/langsci/209)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=209](http://paperhive.org/documents/remote?type=langsci&id=209)

Cover and concept of design: Ulrike Harbort

Typesetting: Change typesetter in localmetadata.tex

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>TeX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

# Contents

<b>1 Interpreting and technology: The upcoming technological turn</b> Claudio Fantinuoli	<b>1</b>
<b>2 Simultaneous interpretation of numbers and the impact of technological support</b> Bart Desmet, Mieke Vandierendonck & Bart Defrancq	<b>3</b>
<b>3 An exploratory study on CAI tools in simultaneous interpreting: Theoretical framework and stimulus validation</b> Bianca Prandi	<b>17</b>
<b>4 Experimenting with computer-assisted interpreter training tools for the development of self-assessment skills: National Parliament of RSA</b> Elizabeth Deysel & Harold Lesch	<b>47</b>
<b>5 Technologies and role-space: How videoconference interpreting affects the court interpreter's perception of her role</b> Jerome Devaux	<b>75</b>
<b>6 Presence? Remote? Remotely present! New technological approaches to remote simultaneous conference interpreting</b> Klaus Ziegler & Sebastiano Gigliobianco	<b>97</b>
<b>Index</b>	<b>117</b>



## Chapter 1

# Interpreting and technology: The upcoming technological turn

Claudio Fantinuoli

University of Mainz

Put abstract here with \abstract.

## 1 Introduction







## Chapter 2

# Simultaneous interpretation of numbers and the impact of technological support

Bart Desmet

Please add your affiliation

Mieke Vandierendonck

Please add your affiliation

Bart Defrancq

Please add your affiliation

In simultaneous interpretation, numbers are a common source of errors. They are often characterized by low predictability from the context and high information density, and the interpreter is therefore required to change strategies with respect to listening, memory and production. Booth technology that automatically recognizes numbers in the source speech and presents them on a screen could reduce the cognitive load and improve translation quality.

In this chapter, we present an experimental study on the properties of numbers that make numbers more or less challenging for the interpreter, and how a technological support system influences performance.

## 1 Introduction

Translation and interpreting are often called sister disciplines, but the integration of the respective activities with technology could not be more different. Computer-assisted translation is now standard practice and machine translation becomes ever more successful to the extent that it seems plausible to some that translators will devote most of their time to post-editing in the near future. In



interpreting, by contrast, technological support is scarce, except for electronic devices used for terminology support in the booth (Fantinuoli2012) or for taking notes in consecutive interpreting and hybrid modes (Orlando2014; Orlando2016; Goldsmith2017). There are a good number of reasons for this discrepancy. First, in the area of interpreting, technology and, in particular, natural language technologies are far less developed than in the area of translation. In fact, automated interpreting systems that are claimed to be effective (such as the Google Pixel Buds) first transform spoken language into some form of written or digital code before translating it and converting it back to a spoken form. Automatic interpreting therefore depends on advances made on the translation front and on the availability of accurate speech-to-text and text-to-speech software. Second, it is well known that spoken language does not come in nicely packed grammatical sentences but is rife with hesitations, unfinished sequences, repairs, etc. which are much harder for an automatic translation system to handle than for a human brain. Replacing the human interpreter with a reliable automatic one will require additional progress in the analysis of human language in context. Finally, even technology-supported human interpreting is slow in developing, as there is little agreement among scholars on the question whether additional sources of information in the booth are really helpful or rather distract the interpreter. There is also evidence of a certain aversion to technology among interpreters (CorpasPastor2016), which is likely to delay the adoption and use of technological support for some time.

The slow progress of technology in interpreting is not for lack of challenges. Simultaneous interpreting is a cognitively demanding task consisting of a variety of processing tasks which have to be carried out in parallel (Gile1995; Seeber2011a) ■ Some sub-tasks are felt by most interpreters to be particularly challenging: the interpretation of numbers, of names, of enumerations, etc. (Gile2009). This paper will focus on numbers and on the effects of (simulated) technological support for the interpretation of numbers. The main research question is whether displaying numbers on a screen in the conference room, immediately after they are articulated by the speaker, increases the accuracy of numbers in the target text. This experimental pilot study thus aims to determine if limited technological support, which would consist of automatic number recognition in the source text and the display of a numerical transcription, is at all helpful in interpreting.

## 2 Numbers in interpreting

There is a very broad consensus among interpreters that numbers are particularly difficult to interpret. Yet, research on the topic of interpreting numbers is rather limited (Mead2015). Starting with Braun1996, several experimental studies have been conducted on the success rate of number rendition in interpreting, showing that overall performance is relatively poor, both in professionals and in student interpreters. In Braun1996, for instance, 12 students obtained a mean error score of close to 70% while simultaneously interpreting numbers contained in speeches. Mazza2001 reports slightly better performances by 15 students, but the mean error rate in her study is still in the 45–50% range. The findings of Pinochi2009, based on interpretations by 16 students, are fairly consistent with Mazza's result. Pinochi compares interpretations from two different source languages (English and German) into one target language and finds that error rates are nearly identical (app. 40%), even though interpreters have to overcome additional challenges due to differences in the syntax of numbers between German and Italian.

Korpal2016 compares student performances with performances by professional interpreters for slow and fast delivery rates. He finds that, although professionals obtain lower error rates than students, nearly 30% of numbers in the interpretation are either wrong or absent altogether in the slower delivery rate. The error rate jumps to 43% for the highest delivery rate. Among students, error rates are in line with Mazza's findings: 44% for the slow delivery rate and 56% for the fast one. Timarová's experimental study of 28 professional interpreters yields an error rate for numbers of approximately 40% (Timarova2012). In a corpus study reported in Collard2017, error rates for numbers in interpretations collected in the European Parliament are close to 18%. One possible – but unverifiable – explanation for the discrepancy between the experimental and the corpus results is the presence of a booth colleague who takes down the numbers for the interpreter ensuring the turn. The numbers can then simply be read off the notebook. In an experimental setting, interpreters perform in isolation, and even though they are usually allowed to take down numbers themselves, this is obviously more difficult than when assistance is provided by a colleague in the booth. In any case, even in naturalistic data, close to one out of five numbers of the source text is rendered incorrectly or omitted in the target language. Based on this, technological support could be helpful in the booth, and it has the potential to reduce the number of errors and omissions.

Scholars have identified several reasons why interpreting numbers is a chal-

lenge. Numbers lack a conceptual representation (Timarova2012; Seeber2015) and are, therefore, not embedded in a conceptual structure allowing interpreters to anticipate them. This lack of predictability of numbers is widely recognised as an obstacle to their interpretation (Braun1996; Mazza2001; Pinochi2009; Mead2015)■ Numbers are highly informative (Alessandrini1990), as every component of a number is a meaningful unit representing only one particular meaning. This prevents interpreters from using strategies such as paraphrasing or reformulation (Pinochi2009). Numbers also usually lack redundant material, which makes them more informative (Gile1995; Seeber2015). Source texts with high information density are known to increase cognitive load in interpreters. As hypothesized by Pinochi2009, differences in number syntax between source and target language can exacerbate this load, as interpreters not only have to render each numeral unit correctly but also swap the order of some of the units (e.g. between English and German).

One of the universally recommended strategies is note-taking (Setton1999; Jones2002; Mead2015): interpreters are advised to stop delivery of the target text as soon as they hear a number, write down that number on a notepad in the booth and read it off while starting up delivery again. The findings of Mazza2001 seem to support the hypothesis that interpreting is more accurate in cases where interpreters jotted down the number. Without a notepad, shortening the ear-voice span (EVS) and changing the listening strategy seem to be the most effective coping strategies: Setton1999 for instance, observes that errors typically occur when the EVS is more than 3–4 seconds. Following Seleskovitch1975, Pinochi2009 advocates a switch from intelligent hearing (i.e. taking into account the context to draw inferences) to literal hearing (i.e. paying attention to the item in isolation).

Assistance by the booth colleague in writing down numbers and visual input provided by the speakers (such as a copy of the speech to be used in the booth, or the projected presentation slides) are said to be beneficial (Mead2015). Lamberger-Felber2001 reports a significant increase of number and name accuracy (53% to 68% fewer errors, depending on the source speech) in an experiment when interpreters are provided the text of the speech in the booth, compared to when they do not have the text at their disposal. Accuracy is highest when, in addition, they are given time to prepare the text they are supplied with. It is to be noted, however, that this is the combined accuracy for numbers and names together. Even for the condition without text, high accuracy rates are obtained (mean of 85.7%). This seems to suggest that names cause significantly fewer problems in simultaneous interpreting than numbers.

While assistance and visual input are likely to boost performance, they are

beyond the control of the interpreter. Technological support could solve that problem: if interpreters could rely on technology that systematically displays numbers as they are pronounced, it could improve the accuracy of the numbers they deliver. Currently, limited applications exist in conference rooms with voting systems, where the results of votes are displayed on a screen in the booth, but the targeted use of natural language processing applications could make it possible in the near future to extract numerical information from online speeches.

### 3 Technology in interpreting

Technology has always been essential to simultaneous interpreting, with audio equipment and booth consoles providing the communication backbone for it to occur. More recently, remote interpreting has been making forays into the profession. **Fantinuoli2018** categorizes these technologies as primarily setting-oriented, since they determine the external conditions in which interpreting takes place. Process-oriented technologies, on the other hand, are designed to support the interpreter in the various phases and processes of interpreting itself, e.g. for the acquisition, organization and retrieval of information, both before and during an assignment. Such technologies aim to directly influence the interpreting process, its associated cognitive load and the quality of its outcome. As such, they are the defining components of computer-assisted interpreting (CAI).

Currently, existing CAI tools are mostly focused on terminological support, whether in preparing for an assignment or for access in the booth. This focus is not surprising – domain-specific terms are an important obstacle to interpreting quality, and CAI tools have the potential of helping interpreters use them more accurately and consistently. Recent studies (**Will2015**; **Fantinuoli2017a**; **Costa2018**) have surveyed existing CAI tools for terminology management, and determined relevant criteria to evaluate them. For a knowledge management tool to be practical in the booth, one important requirement is that it allows the interpreter to access reference material quickly and with as little additional cognitive load as possible. This can be achieved with good knowledge representation, clear presentation and ergonomic operation, and good search algorithms.

It is essential that relevant information can be retrieved fast, i.e. within the ear-voice span. Automatic Speech Recognition (ASR) has the potential of speeding up the look-up process and solving the cognitive effort and latency of manual querying. Technological advances in the field have been rapid in recent years, especially since the introduction of neural networks to solve the task (**Yu2016**). Given the current state of the art in ASR and its foreseeable progress, it seems to

be a matter of time before the technology is used in CAI tools to support interpreters with terminology look-up, and/or with information-dense content like numbers, as explored in [Fantinuoli2017b](#). This paper will focus on the latter of these aspects.

While the development and adoption of CAI tools has been limited, scientific research on the impact of their use has been even scarcer. The main contribution of this work is that it experimentally evaluates the potential impact of ASR-driven CAI support that displays numbers on-screen in real time.

## 4 Experiment

The aim of this experiment is twofold: to determine if limited technological support can improve the accuracy of interpreted numbers, and how this improvement breaks down over different number and error types. In the following, we describe the system used as a proxy for automatic number support, the participants, selected speeches and the distribution of numbers in them, the experimental setup and the parameters for evaluation.

### 4.1 A proxy system for automatic number support

An ideal system for number support during simultaneous interpretation would consist of three components:

1. ASR that can transform an incoming speech signal into text quickly, accurately and without the need for being tuned to a specific speaker or accent,
2. software to isolate numbers from the text in meaningful units, and
3. a way of ergonomically presenting those numbers to the interpreter.

Since no such system existed at the time this study was conducted, a mock-up system was used to simulate the desired behavior. Microsoft PowerPoint presentations were prepared ahead of time based on the speech transcripts, containing one slide per number in the speech. Figure 1 shows an example slide as used in the experiment. Numbers were presented in numerical format with spacing between multiples of a thousand, and formatted in a large fixed-width font. The two previous numbers (if available) were displayed above the focus number, so that numbers in rapid succession would stay accessible longer. During the experiment, the presentation was shown on a big screen in the conference room,

Converted  
from  
inline  
enumer-  
ation to  
list

and slides were manually advanced immediately after a number had been fully pronounced, i.e. simulating an automatic system with minimal latency.

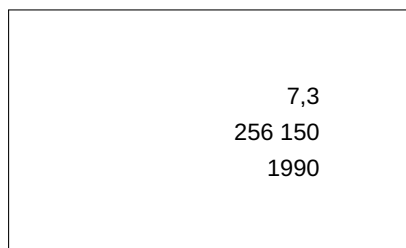


Figure 1: Example of the mock-up technological support, showing the three most recent numbers that have been pronounced by the speaker (Microsoft PowerPoint slide). New numbers are added to the bottom of the slide, with old numbers shifting upwards. A maximum of three numbers is shown.

### 4.2 Participants

The experiment was performed with ten interpretation students enrolled in the postgraduate program for Conference Interpreting at Ghent University with Dutch as an A-language. Seven of them had completed a 4-year applied linguistics program with a focus on French, the three others with a focus on German, making them proficient at the C2 level according to the CEFR framework (Council2001). At the time of the experiment, all participants had received 5 weeks of simultaneous training at the postgraduate level and were taking additional *retour* classes for the source languages of the experiment. They were all graduates of a master program in interpreting that offers an introduction to simultaneous interpreting and limited practice. Participants were aged between 22 and 27 years and 9 of them were female.

### 4.3 Speeches

Four experimental speeches of similar difficulty and length were prepared, with parallel versions in French and German. The average text length for French was 1121 words, the German texts conveyed the same content with almost 10% fewer words, at an average of 1022 per text. German compounds are written as a single word, which largely explains this discrepancy. The texts dealt with diverse topics (Amazon, child labour, inheritance law and natural disasters), and specific terminology was provided to the participants before the start of each speech.

Like in the experiments of **Braun1996** and **Mazza2001**, the start of each speech (150 words) contained no numbers, to allow participants to settle in. After that, 20 numbers occurred at random intervals. They were equally distributed over four categories, with each text containing exactly 5 instances of each category: simple whole numbers (e.g. 87 or 60 000), complex whole numbers (e.g. 387 or 65 400), decimals (e.g. 28.3) and years (e.g. 2012). We distinguished between simple and complex whole numbers based on the number of meaningful units rather than the number of digits, since large numbers can be conceptually simple (e.g. 1 million contains only 2 meaningful units). In this study, numbers containing 3 or more meaningful units were considered complex.

#### 4.4 Setup

Participants were required to interpret into their A-language: Dutch. The experiment was conducted in two sessions: one for French with 7 participants, and one for German with 3 participants. There was no overlap between the two participant groups. Offering two source languages created an opportunity to check the results for the influence of number syntax: German and Dutch are both “unit-decade” languages (61 is *eenenzestig* in Dutch: ‘one-and-sixty’), whereas French is a “decade-unit” language (*soixante-et-un* ‘sixty and one’. In interpreting from French into Dutch, the order of certain units inside the number needs to be changed, whereas if German is the source language, no such changes are required.

Before the experiment, participants were informed that numbers would occur with high frequency in the speeches, and they were familiarized with the mock-up technological support. Each speech was then read by a near-native, and recorded digitally, along with all interpretations into Dutch. The first two texts were interpreted without simulated technological support, whereas it was available for the last two. The rate of delivery, a potential problem trigger for interpretation quality (**Gile1995**), was within the optimal range for interpreters: an average of 121 words per minute for French, and 101 words per minute for German. This difference in rate can be partly explained by the higher information density of German words, but the German speaker also took more time to deliver speeches with identical content to their French counterparts: an average of 608 seconds, 9.4% longer than the 556 seconds for French. Within the same language, the delivery rate for speeches interpreted without technological support was slightly slower than those interpreted with support: an average of 115 and 128 words per minute for French, and 100 and 102 words per minute for German, respectively. Any benefits from the technological support can therefore not be



attributed to a slower rate of delivery.

After the experiment, a questionnaire polled whether the participants had found the technological support helpful or distracting, how long it took to get used to it, and what they would change about it.

## 4.5 Evaluation

Given the focus of this study on interpreting numbers, the recordings were analyzed to produce systematic records for each number. These include a transcript of the number stimulus in the source speech and the provided interpretation, which allows the performance to be coded: has the number been interpreted correctly, and if not, what kind of error was made? For the categorization of errors, we follow **Pinochi2009**, which in turn was adapted from **Braun1996**:

**Omission:** the number is missing or has been replaced by a general expression (e.g. 47 becomes *many*).

**Approximation:** the order of magnitude is correct, but the number has been rounded up or down (e.g. 47 becomes *around 50*). These adaptations can be viewed as interpretation tactics rather than mistakes.

**Lexical mistakes:** the order of magnitude of the number is correct, but some of its components have been changed (e.g. 47 becomes 49).

**Transposition:** all components are correct, but their order has been changed (e.g. 47 becomes 74). This error can be especially frequent when the source and target language use a different number syntax (pronouncing units or tens first).

**Syntactic mistakes:** the order of magnitude is incorrect, even though the right components may be present (e.g. 47 becomes 470).

**Phonological mistakes:** the error can be explained by phonological confusion in the source stimulus (e.g. 14 becomes 40, a near-homophone in English).

**Other mistakes:** miscellaneous errors that do not fit any of the other categories, or numbers that combine multiple error types (e.g. 47 becomes 740).

Additionally, the records contain information on the source language, source speech, participant, the type of number, and whether the experimental technological support was available (i.e., the independent variable). These variables allow the impact of CAI support to be analyzed from multiple perspectives, as presented in the following section.

## 5 Results and discussion

The experiment yielded 40 recordings of 4 speeches interpreted by 10 participants. French was the source language in 28 of them, German in 12, and half of them were interpreted with technological support for numbers. With 20 numbers per recording, a total of 800 observations was available for this pilot study.

### 5.1 The impact of technological support for interpreting numbers

Numbers are difficult to interpret, and this is borne out by the experimental results. The average accuracy on numbers in the control setting, without technological support, was 56.5% over all test subjects and languages, i.e. an average error rate of 43.5%. The observed performances were thus completely in line with the 40–50% error rate found in most other studies. Accuracy between individuals varied from 27.5% to 90%, and the interpretations from German were more accurate on average (74% vs 49%). This could be due to the higher similarity between the German and Dutch number systems, but the number of participants is too small to rule out individual differences as the main cause of this divergence. Another explanation is the slower rate of delivery for German than for French.

When technological support was made available, the performance on numbers improved dramatically. Average accuracy rose by 30 percentage points to 86.5%, and the absolute number of errors dropped from 174 to 54 out of 400, an error reduction of 69%. This finding was fairly consistent with the 53–68% decrease reported by **Lamberger-Felber2001** for numbers and names, when copies of speeches were made available to the interpreter. A paired t-test showed that the performance difference was highly significant for both source languages ( $p < 0.001$ ). It should not come as a surprise that individuals that scored poorly without support benefit most from having it, at least in absolute terms of avoided errors (up to 21 fewer errors on 40 numbers). Interestingly, relative error reduction was moderate (43%) to high (90%) regardless of a subject's performance in the control setting. In other words, even if an interpreter was highly competent at conveying numbers without support, he or she was able to reduce their error rate significantly when support is available.

It is obvious from these results that CAI support for numbers has the potential of drastically reducing errors on numbers. Nevertheless, it needs to be emphasized that these results indicate the ceiling performance of what can be achieved with technological support – or with a dedicated booth mate, for that matter. The tested mock-up system has minimal delay and its output is entirely accurate. With such a reliable system, the interpreter can choose to reduce the listening

and memory efforts to focus more on production. Further research is required to test how imperfect output would affect the distribution of effort.

Feedback from the questionnaires indicated that all participants perceived the support as helpful. Three participants reported that it took some time to get used to the system, for the others the transition was almost immediate. Two respondents were sometimes distracted by the system because it focused their attention on numbers at the expense of other content. Future evaluations should elucidate if distraction can be reduced with increased familiarity with number support, or different ways of providing it. Specifically, adding units (e.g. *km* or *percent*) or concepts to the numbers could be beneficial to recall their usage in context, although this may require cues that are specific to the source language.

## 5.2 The influence of number type

The numbers expressed in the experimental speeches were balanced over four categories: simple and complex whole numbers, decimals and dates. Figure 2 shows the performance in terms of accuracy, categorized by these number types.

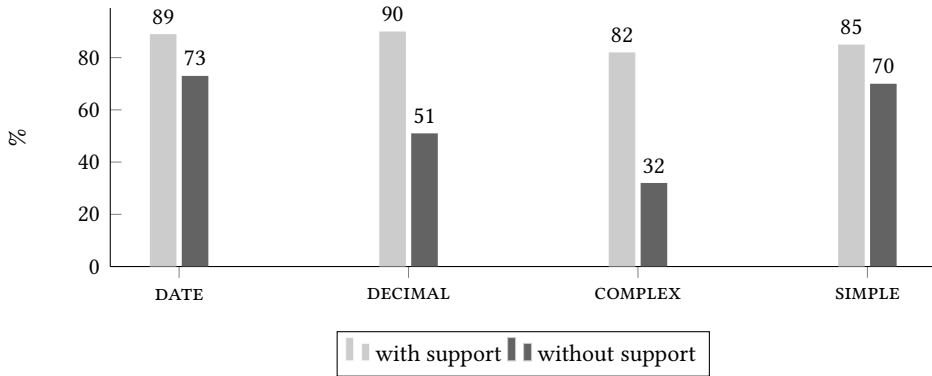


Figure 2: Average accuracy per number type in both experimental conditions

It can be observed that in the standard setting without support, interpreters experience most difficulties with complex and decimal numbers. Accuracies for these types are low at 32% and 51%, respectively, compared to around 70% for both simple numbers and dates. These results corroborate previous findings. This study used the same number typology as Mazza2001, with the exception that we distinguish whole numbers based on complexity rather than size; Pinochi2009 separates whole numbers into three categories, based on the complexity of pro-

nouncing them. In all three studies, large or complex whole numbers are found to cause the most errors, followed by decimals, and small or simple whole numbers and dates are the easiest to interpret.

When interpreters have access to technological support, the differences in error rate per number type almost disappear. Overall, we see that each type benefits from it. As expected, the largest gain is for complex whole and decimal numbers, with absolute error reductions of 50 and 39%, respectively (significant at  $p < 0.01$ ). Even for the simpler number categories, however, there are significant improvements ( $p < 0.05$ ). With technological support, accuracy is almost identical across number types, which suggests that the remaining errors are due to factors other than the complexity of the number. A detailed analysis of the remaining errors is still to be made.

### 5.3 Error type analysis

A total of 228 errors was observed in the experiment. Figure 3 presents them separated by type.

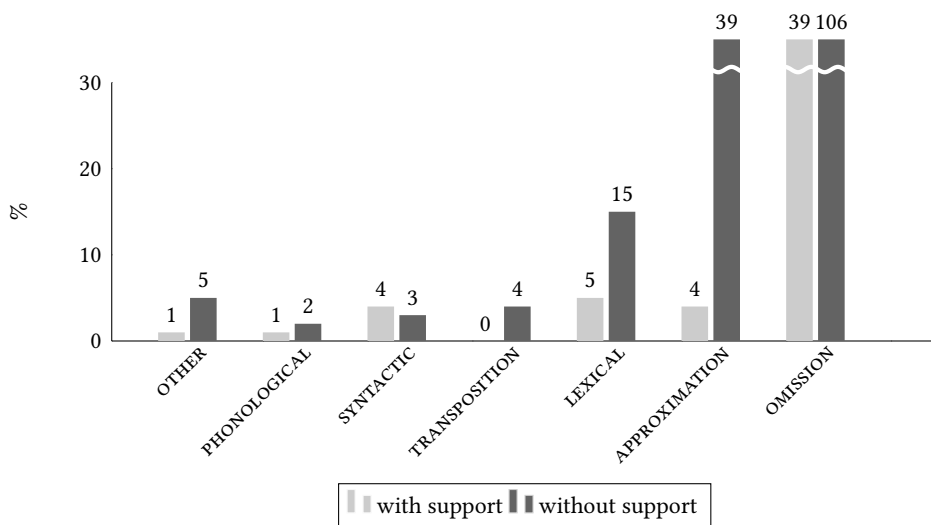


Figure 3: Error count per type in both experimental conditions

By far the most frequently occurring error is omission, regardless of the availability of technological support. Technological support significantly reduces the number of omissions ( $p < 0.05$ ), but omissions remain frequent and their relative weight in the total error load increases from 61 to 72%. We see two possible expla-

nations for this. Some numbers come in information-dense sections of a speech, and omitting them may be a necessity to limit cognitive load or to avoid increasing EVs. In such cases, even having the number available on screen would not help in reproducing it in context. Second, it could be that technological support causes complacency or confusion when numbers are heard. If an interpreter reduces the listening effort and relies on a support system to receive the number, adequate context may be lacking to convey its meaning.

Approximations are often used in the control setting, but only when the source stimulus was a decimal or a complex whole number. Approximating is a useful strategy when numbers have not been entirely understood, or when they are too large to fit in working memory. With the addition of technological support, this strategy is used almost 10 times less often, since the tasks of comprehension and working memory are effectively solved. The error reduction is significant at  $p < 0.001$ .

Lexical mistakes, the third most frequent error in the control setting, occur three times less often when support is available. The other four error categories do not occur frequently in either of the settings. Differences between the two experimental settings are not significant or lack support.

The error distributions are in line with the findings of Mazza2001 and Pinochi2009 who also found omissions and approximations to be most frequent, in that order.

## 6 Conclusions and future work

This paper presented an experimental pilot study of the potential impact of booth technology that supports the interpretation of numbers. Our mock-up system simulates technology that automatically recognizes numbers in the source speech and presents them on a screen in the conference room, in order to reduce the cognitive load and improve translation quality.

Technological support improves overall accuracy on numbers from 56.5 to 86.5 percent, reducing the amount of errors by two thirds. The improvement is statistically significant for all participants. Technological help is most helpful in reducing errors on complex numbers and decimals, the two categories that are most often interpreted incorrectly. Omissions are the most frequent error, followed by approximations. The occurrence of the latter drops by almost 90 percent when support is available.

Since the experiment was performed with students, the results are not readily applicable to professional interpreters. Even though the outcome of the experiment clearly shows the potential of CAI support for numbers, it must be em-

phasized that our experimental design is not based on automatic recognition of numbers in speech. Automatic Speech Recognition might not achieve perfect recognition and minimal latency. Therefore, our results describe the ceiling performance that could be achieved with such an ideal system. Further studies should be carried out on how interpreters deal with discrepancies between auditory input from a speaker and visual input from an automatic recognition system, increased delay or different modes of presentation. Further research should also focus on the rendition of items used in combination with numbers, as it is known that interpreters tend to direct so many of their attentional resources to numbers that errors also frequently occur in the context of numbers (Gile2009).

## Chapter 3

# An exploratory study on CAI tools in simultaneous interpreting: Theoretical framework and stimulus validation

Bianca Prandi

University of Mainz

The acquisition of terminology and specialized knowledge prior to a technical conference represents a fundamental phase in the interpreter's workflow, but quick and easy access to terminological information during the interpreting task is equally important to support the interpreter in the rendition of terminology and to ensure a high-quality interpreting performance. The processing of terminology during simultaneous interpreting represents a sub-task that can be a cause of saturation and lead to errors or omissions. In order to facilitate the processing of terminology and avoid cognitive overload, interpreters resort to various terminology management solutions.

Over the past few years, terminology management tools have been developed specifically for interpreters. Known as Computer-Aided Interpreting (CAI) tools, they are designed to support interpreters in their entire workflow while keeping the additional cognitive load deriving from the terminology search during interpreting as low as possible. The role of CAI tools is now being acknowledged by practitioners and interpreting scholars, but the impact of such tools on the cognitive processes involved in simultaneous interpreting is still unclear. A project underway at the University of Mainz/Germersheim aims at bridging this knowledge gap by adopting an empirical approach.

To this end, an exploratory study was conducted to evaluate the appropriateness of the stimuli adopted for data collection and to verify whether the use of CAI tools causes saturation or, on the contrary, helps prevent it by reducing the local cognitive load during terminology search and delivery of the target text. In particular, the terminology search is expected to be quicker and more effective than for other solutions and the added cognitive load to be lower. In this paper, after describing



the theoretical framework adopted in the study and our hypotheses, we present the structure of the pilot study and the stimuli used. We then validate the stimuli used and present the first results derived from the analysis of the test subjects' renditions.

Abstract needs to be shorten, such that the main text can begin on the title page.

## 1 Introduction

Computer-Assisted Interpreting emerged around 10 years ago to provide interpreters with tools to prepare for specialized events and to support them along the individual phases of their workflow, from preparation, to interpretation proper, to follow-up work after the assignment. CAI tools thus rationalize the interpreter's terminology work by making preparation more efficient and ultimately aim at improving the quality of the interpreter's output, at least in terms of terminological precision and adequacy. **Rütten2007** and **Will2009** developed a theoretical model of the interpreter's preparation work and laid the foundations for how a CAI tool should be structured in order to address the specific needs of conference interpreters, which are mainly linked to the online nature of interpreting and the time constraints it entails.

To date, the number of CAI tools available to interpreters is limited and their functionalities do not always cover all the phases of the interpreting process. **Fantinuoli2018** distinguishes between first-generation and second-generation CAI tools. The first (e.g. *Interplex*<sup>1</sup> and *Terminus*<sup>2</sup>) are “designed to manage multilingual glossaries in an interpreter-friendly manner” (**Fantinuoli2018**), but do not offer an advanced search algorithm. The latter “offer advanced functionalities that go beyond basic terminology management, such as features to organize textual material, retrieve information from corpora or other resources, learn conceptualized domains, and advanced search functions” (**Fantinuoli2018**) and include *Intragloss*<sup>3</sup> and *InterpretBank*<sup>4</sup>. *Interpreter's Help*<sup>5</sup> can also be considered a second-generation CAI tool, as it implements an advanced search function through its companion tool *Boothmate*<sup>6</sup>.

---

<sup>1</sup><http://www.fourwillows.com/interplex.html>

<sup>2</sup><http://www.wintringham.ch/cgi/ayawp.pl?T=terminus>

<sup>3</sup><http://intragloss.com>

<sup>4</sup><http://www.interpretbank.com>

<sup>5</sup><https://interpretershelp.com>

<sup>6</sup><https://interpretershelp.com/boothmate>



Given the relatively recent introduction of these tools on the market and their varied nature, first attempts at an evaluation have been made. Two main trends can be identified in this respect. The most recent one focuses on developing a set of criteria against which the tools can be evaluated (Costa2016; Will2015). This approach is certainly ambitious, but it remains somewhat arbitrary. The evaluation criteria mainly reflect the features offered by the tools, but do not consider how they influence the product of the interpreting process in terms of terminological quality and whether they optimize the interpreters' preparation and facilitate their work in the booth, by making the online retrieval of terminological units easier and improving the terminological quality. While the number and type of features of CAI tools certainly is of interest for practitioners, we would argue that the main reason for choosing to use CAI tools, and to prefer one tool to the other ones available, should be the ability of such tools to positively influence the interpreter's work in terms of cognitive capacity and, ultimately, quality.

This is where the second trend in the evaluation of CAI tools comes into play. Soon after the development of the first CAI tools, the first studies on the topic appeared. Apart from a few Master's theses, which are limited in scope and often take a descriptive approach, rather than an investigative one (see for example DeMerulis2013), a rather small number of publications can be found which mostly deal with the application of CAI tools to the preparation phase (Xu2015; Fantinuoli2017a). When it comes to the use of CAI tools in the booth, the number of studies is very limited, as is their scope. First attempts at an empirical analysis of the use of CAI tools during SI can be identified in Prandi2015a; Prandi2015b and Biagini Biagini2015. These initial investigations of the issue speak in favor of the usability of CAI tools and seem to suggest that they do improve terminological quality of simultaneous interpreting (SI). Both experiments were based on a product-oriented analysis of the test subjects' deliveries. Biagini also included a statistical analysis of transcription data. Apart from these initial analyses, no empirical methodology has been tested in a wide-ranging experiment which implements psycho-physiological, process-oriented methods in addition to product-based analysis. In addition, Fantinuoli2017b recently addressed the topic of the integration of automatic speech recognition (ASR) in CAI tools for the use in the booth.

A PhD research project underway at the Johannes Gutenberg University of Mainz/Germersheim (Prandi2016; Prandi2017a; Prandi2017b) aims at bridging this research gap. By triangulating pupillometry data, eye-tracking data and the analysis of the test subjects' transcriptions, we aim at providing a picture not only of the usability of CAI tools during simultaneous interpreting, but also of the

local variations in Cognitive Load (CL) and of the terminological quality of simultaneous interpretation performed with the support of a CAI tool when compared to more traditional terminology management solutions. Through our study, we hope to develop a research methodology that can be used to evaluate CAI tools and provide the much-needed empirical data that will be helpful not only to practitioners in choosing the best tool, but also to software developers, by highlighting potential shortcomings.

After discussing the theoretical framework of our analysis (§2), we present our research desiderata and the structure of an exploratory study conducted to test our research methodology (§3). §4 describes the rationale behind the stimuli used in the experiment and the features of the speeches used. We then present the results of the analysis of the transcriptions (§5), which we use to evaluate the appropriateness of our stimuli. In the conclusions, we address future work and provide suggestions for further research.

## **2 Theoretical framework**

In our investigation of simultaneous interpreting performed with the support of CAI tools, our aim is not only to look at the product of such activity, but also at the process that lies behind it. For this reason, in establishing a theoretical framework for our analysis, we took into consideration the two theoretical models that set out to describe interpreting from a procedural point of view and that address the allocation of cognitive sources during this very complex mental activity: Gile's Effort Model (EM) of SI and Seeber's Cognitive Load Model (CLM) of Simultaneous Interpreting. In this section, we discuss why Seeber's approach is more suited to the operationalization of our hypotheses.

### **2.1 Gile's Effort Model and Seeber's Cognitive Load Model of Simultaneous Interpreting**

The main point of divergence between Gile's Effort Model of Simultaneous Interpreting (Gile1988; Gile1997; Gile1999) and Seeber's Cognitive Load Model of SI lies in the theoretical assumptions they stem from. Gile draws from Kahneman's single resource theory (Kahneman1973), which does not find much validation in scientific literature. If there is one single pool of resources interpreters can adopt, how can some interpreters perform a terminological search on the Internet, while at the same time delivering a perfectly acceptable rendition of the original speech? This kind of multi-tasking might seem impossible to a first-year

interpreter trainee, but is commonly observed among experienced professional interpreters. The second controversial assumption is that interpreters work close to saturation level most of the time (Gile1999's "tightrope hypothesis", Gile1999). While this might be true in some cases, for instance when the source speech is particularly dense, fast, or pronounced with a non-native accent, there might very well be cases in which the interpreter has enough spare cognitive resources to do something else while interpreting.

In his Cognitive Load Model of Simultaneous Interpreting, Seeber takes an opposite approach to Gile's, basing his model on Wickens's multiple resource theory and on his Cognitive Load Model (Wickens1984; Wickens2002). Wickens developed his model to account for the fact that qualitative differences in tasks being performed at the same time lead to "differences in time-sharing efficiency" (Wickens2002), as shown by Kantowitz1976 and Wickens himself (Wickens1976). According to Wickens, different kinds of tasks require resources that are managed by discrete structures. When two or more tasks are performed simultaneously and "all other things [are] equal (i.e. equal resource demand or single task difficulty), two tasks that both demand one level of a given dimension (e.g. two tasks demanding visual perception) will interfere with each other more than two tasks that demand separate levels on the dimension (e.g. one visual, one auditory task)" (Wickens2002). In other words, performing a visual and an auditory task simultaneously will be "easier" (i.e. more efficient), because the underlying structures are not shared, than performing two visual tasks, as they share the same structures. In his model, Wickens identifies four dimensions, each made up of two "levels":

- processing stages (perception & cognition<sup>7</sup>/responding)
- processing codes (spatial/verbal)
- processing modalities (visual/auditory)
- visual processing (ambient/focal)

Not shown in the graphic representation of the model (Figure 1), but postulated by Wickens, is an additional pool of general capacity, which is always available to all tasks. In his adaptation of Wickens's model to simultaneous interpreting, Seeber takes a step further, simplifying the graphical representation of Wickens's

---

<sup>7</sup>Perception and cognition are considered as one dimension, as one cannot take place without the other.

model by turning it into a 2D model (see Figure 2). This has two main advantages. First, it allows seeing all “sides” of the cube (i.e. all dimensions) at once. Second, it graphically introduces the general capacity left out in Wickens’s “cube”.

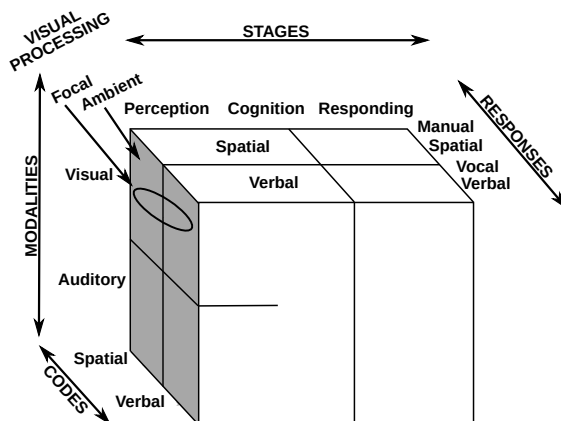


Figure 1: Cognitive Load Model (adapted from Wickens2002)

The result of this adaptation is a Cognitive Resource Footprint (CRF), which Seeber also develops for shadowing and sight-translation (Seeber2007). Simultaneous interpreting is the combination of two main tasks: the listening and comprehension task on the one hand, and the production and monitoring task on the other. As shown by Figure 2, the first task mobilizes auditory-verbal and cognitive verbal resources at the perceptual-cognitive stage (interpreters receive the aural stimulus, i.e. the words pronounced by the speaker, and analyze the verbal message). The second task requires the same kind of resources at the perceptual-cognitive stage and additional vocal-verbal resources at the response stage (interpreters verbally “respond” to what they have heard by delivering the message in the target language, but also listen to and monitor their own rendition).

The footprint is integrated by a Conflict Matrix, which shows the degree of interference between two co-occurring tasks as the sum of the demand vectors for each sub-task and of the individual conflict coefficients between sub-tasks (see Figure 3).

The demand vectors indicate the degree to which each sub-task recruits a certain type of resource. Seeber postulates a demand vector of 1 for each sub-task. Conflict coefficients instead show to which degree the single sub-tasks compete for the same resources. When two sub-tasks share resources that are governed by the same structures, their level of conflict is higher than for two sub-tasks that do

### 3 An exploratory study on CAI tools in simultaneous interpreting

**Simultaneous interpreting [total interference score: 9]**

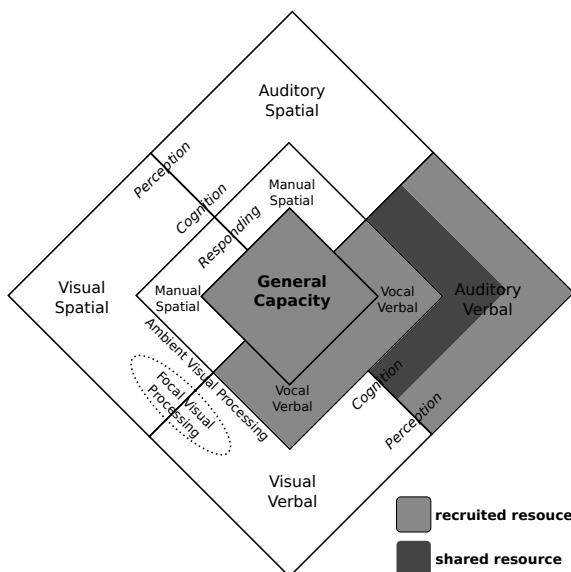


Figure 2: Cognitive resource footprint for simultaneous interpreting (adapted from Seeber2007)

not share resources (and time-sharing between them is not as efficient). The sum of demand vectors and conflict coefficients gives a value of 9 for simultaneous interpreting.

The possibility to “quantify” the degree of interference between co-occurring tasks and to explain multi-tasking makes Seeber’s Cognitive Load Model more suited than Gile’s EM to formulate hypotheses on Simultaneous Interpreting with CAI tools, as discussed below. For this reason, we chose the Cognitive Load Model for simultaneous interpreting as our theoretical framework.

## 2.2 Hypotheses on SI with CAI

Seeber uses his model to represent the allocation of cognitive resources during “standard” simultaneous interpreting, without indicating any specific conditions under which this activity is performed. What happens when, during SI, the interpreter can query a terminological database? What kind of cognitive resources are recruited, and at which stage? And how much do they interfere with each other?

In addition to the operations traditionally performed during simultaneous in-

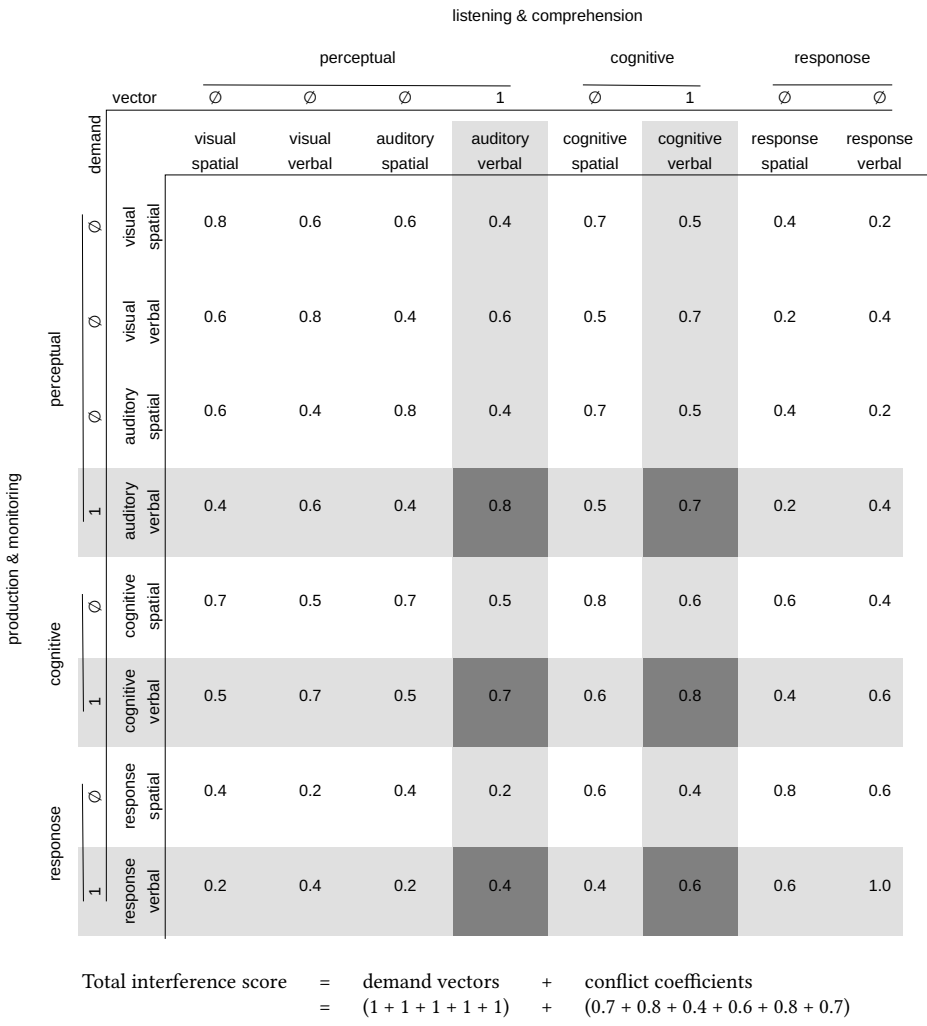


Figure 3: Conflict matrix for simultaneous interpreting (adapted from Seeber2007)

ToDo for Langsci: Spacing between rows and columns, better rules!

interpreting, when working with a CAI tool, or with another terminology management solution – such as an electronic glossary – the interpreter has to type a term or part thereof in order to query the database. This action can be considered as a response to the auditory stimulus, a reaction that precedes the vocal-verbal response (i.e. the interpreter’s delivery of the term in question). During the look-up process, manual-spatial resources are therefore recruited at the response stage. After the query has been completed, the interpreter is typically presented with a list of terminological pairs (the term and its equivalent(s) in the target language). He or she will therefore need to visually identify on the screen the term needed, an operation that requires visual-spatial resources at the perceptive-cognitive stage. Once the term has been identified, it is also read, making use of visual-verbal resources in the same stage of the process. As illustrated by Figures 4 and 5, the Cognitive Resource Footprint for simultaneous interpreting during which a terminological query is performed using a CAI tool or an electronic glossary recruits more resources than “standard” SI.

It goes without saying that the CRF shown in Figure 5 applies only to those moments when the interpreter is performing a query, and should not be seen as representative of the whole interpreting process. Cognitive load is not static, but varies constantly during the interpreting process, as a function of the cognitive resources recruited. We hypothesize that cognitive load is higher while the query is performed, since more cognitive resources are recruited (as shown by the CRF). In some cases, it might even lead to cognitive overload. If the term retrieval is successful, however, we expect cognitive load to go back to normal levels during production. Cognitive load might even be lower than for “standard” simultaneous interpreting, as the search for the appropriate term in the interpreter’s memory would be replaced by a query in the glossary.

If we only took into consideration the Cognitive Resource Footprint, we would not, however, be able to formulate hypotheses on the differences in Cognitive Load experienced while working with a CAI tool or with less advanced terminology management solutions, such as electronic glossaries in the form of a Word or Excel table. These differences can be explored by assigning a higher demand vector to the various terminology management solutions. The Conflict Matrices can thus help visually represent the different levels of recruitment of cognitive resources. If the glossary is the same, what varies among the tools are the user interface and the search algorithm. The most advanced CAI tools, and Interpret-Bank<sup>8</sup> in particular, which we adopt in our study, are designed to yield the most accurate results and to facilitate the user in identifying the term needed on the

---

<sup>8</sup><http://www.interpretbank.com>

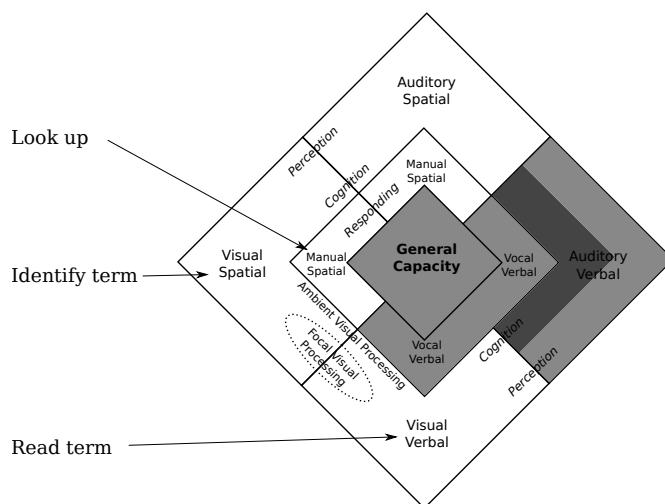


Figure 4: Additional cognitive resources recruited during SI with a CAI tool/electronic glossary

screen. We therefore expect the tools to require a lower level of manual-spatial resources (to look up the term) and of visual-spatial resources (to locate the term on the screen), when compared, for instance, to an Excel spreadsheet. As shown by Figures 7 and 6, we can therefore assign a demand vector of 1 to each of these resources in the case of CAI tools, and a demand vector of 2 in the case of an Excel spreadsheet. The total interference score for SI performed during the use of a CAI tool will therefore be equal to 14.8, while for SI with the use of an Excel spreadsheet it will be higher (16.8).

The integration of automatic speech recognition in a CAI tool (see Fantinuoli2017b) would require no manual-spatial resources, thus lowering the total interference score to at least 13.2.

### 3 Designing a pilot study on the use of CAI tools in the booth

#### 3.1 Introduction

The debate around how CAI tools influence the process and the quality of interpretation is in large measure not based on empirical data, which are still very scarce and limited to a few small experiments, but is rather the result of personal beliefs and assumptions which have not been proven empirically. A research project cur-



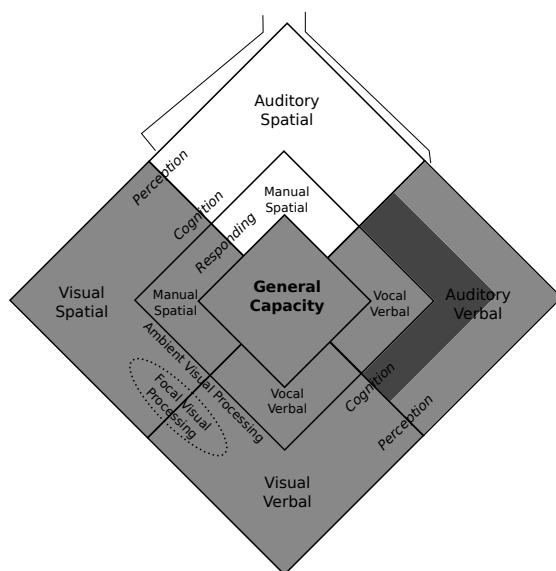


Figure 5: Cognitive Resource Footprint for SI with a CAI tool/electronic glossary

rently underway at the University of Mainz/Germersheim (**Prandi2016; Prandi2017a; Prandi2017b**) aims at bridging this research gap by providing data that can substantiate arguments in favor and against CAI tools. One source of difficulty in the investigation of CAI tools lies in the fact that no research methodology for the combined collection of data both on the process and on the product of SI with CAI has been developed and tested yet. In order to provide a first solution to this issue, we therefore conducted an exploratory study with the aim of evaluating the appropriateness of the stimuli used for data collection. In the following sections we will present our research questions, describe the structure of the exploratory study and illustrate the stimuli used. The analysis of the participants' renditions is the subject of the remainder of the paper.

### 3.2 Research questions

Our research projects aims at answering three fundamental questions:

listening & comprehension									

		listening & comprehension								
		perceptual				cognitive		response		
		vector	∅	∅	∅	1	∅	1	∅	∅
production & monitoring	perceptual	demand	visual spatial	visual verbal	auditory spatial	auditory verbal	cognitive spatial	cognitive verbal	response spatial	response verbal
		∅	0.8	0.6	0.6	0.4	0.7	0.5	0.4	0.2
		∅	0.6	0.8	0.4	0.6	0.5	0.7	0.2	0.4
		∅	0.6	0.4	0.8	0.4	0.7	0.5	0.4	0.2
		1	0.4	0.6	0.4	0.8	0.5	0.7	0.2	0.4
	cognitive	∅	0.7	0.5	0.7	0.5	0.8	0.6	0.6	0.4
		1	0.5	0.7	0.5	0.7	0.6	0.8	0.4	0.6
	response	∅	0.4	0.2	0.4	0.2	0.6	0.4	0.8	0.6
		1	0.2	0.4	0.2	0.4	0.4	0.6	0.6	1.0

Figure 7: Conflict matrix for SI with CAI (InterpretBank) TIC = 14.8

ToDo for Langsci: Spacing between rows and columns, better rules!

- Do CAI tools help improve the terminological quality of the interpretation when compared to traditional electronic glossaries?
- Does a query performed with a CAI tool during SI lead to lower additional local cognitive load when compared to traditional glossary prepared with Word or Excel? Does looking up terminology lead to cognitive overload and if so, does this also happen when CAI tools are used?
- Can a combination of eye-tracking measures, pupillometry and transcription analysis be used to acquire data on the interpreting process, the terminological quality of the product and the usability of CAI tools?

In order to collect first data to help answer these questions, an exploratory study was conducted between May and July 2017 at the University of Mainz/Germersheim. For the scope of this paper, we will report on the observations made during the analysis of the product, while further work will be required to address the issues related to the process of simultaneous interpreting with CAI tools.

### **3.3 Structure of the study: sample, duration, training and data collection**

The pilot study involved 6 advanced students of the Master's degree in conference interpreting of the University of Mainz/Germersheim. Prior to the study, all students had had at least 3 semesters of practice in simultaneous and consecutive interpreting and had English in their combination, as B or C language. Half of the sample was made up of German natives (one male and two females), half of Italian natives (one female and two males). The test subjects were recruited by e-mail and their participation in the experiment was voluntary. No monetary compensation was offered, but the participation in the study gave the trainees the opportunity to learn about a new tool, InterpretBank, and to practice in the booth with a laptop, something they rarely do in class.

The trainees attended one preliminary meeting during which the basics of terminology management for conference interpreters were covered. The presentation was centered on practice rather than theory, since a previous study confirmed this was more beneficial to achieve a good level of expertise (Prandi2015a; Prandi2015b). The search functions in Word, Excel and InterpretBank were described in detail. For the purpose of the study, participants could visualize all the results of a query when working with Word<sup>9</sup>, while they had to skip to the

---

<sup>9</sup>The results are displayed in a column on the left-hand side of the window.

next occurrence when using Excel. In our presentation we made sure to choose a neutral approach to the different tools, so as not to favor the CAI tool chosen.

After covering the basics, 5 practice sessions followed in the subsequent weeks, with around 1 session per week. During each training session, the students interpreted 3 short speeches from English into their mother tongue (either German or Italian). They could use a glossary provided by us, for both language combinations, which they could access in all three formats (Word, Excel or InterpretBank). During each session, they used a different tool for each speech, so equal practice time was dedicated to each tool. The first few speeches had been prepared ad-hoc by the author for a previous study (Prandi2015a; Prandi2015b), while the last few speeches were authentic speeches selected by us, so as to ensure a certain progression in the practice material. The topics covered during the practice sessions were medicine and biology. After the last session, the students took a short test to verify their proficiency in the use of the tools. All students passed the test and were deemed ready for data collection.

Data collection took place in the *Translation and Cognition Centre* of the University of Mainz/Germersheim. The test subjects were briefed about the structure of the study and were informed that they were going to interpret 3 speeches from English into their mother tongue. They were told the topic of the speeches (renewables and other sources of energy) right before data collection started. While this does not reflect professional practice, which requires thorough preparation before interpretation proper, the students were not given the chance to prepare in advance since this would have introduced an additional variable in the study. The methods of preparation and the time dedicated to this fundamental phase of interpreting are very personal and would have been very difficult to standardize and to verify. We therefore decided to sacrifice some ecological validity to limit the number of independent variables.

Every test subject interpreted three speeches, each about 12 minutes long and with an average speed of 122.26 words per minute. This speed was chosen to make sure that looking up terminology during interpreting was challenging, but not impossible. All three speeches had been prepared ad-hoc for the study and previously recorded by a native speaker of British English. One glossary of 421 terms was prepared by us. It contained the same terms for both language combinations and had a simple tabular structure – one column for the source language and one for the target language. The glossary was prepared with InterpretBank and then exported as an Excel spreadsheet, which was then also converted in a Word table. The glossaries were not shown to the test subjects before the interpreting task started. During the interpreting task, the screen was divided in two

areas. On the left-hand side, the test subjects could see the video of the speaker, which served as a fixation cross when no term query was being performed. The glossary window was placed on the right-hand side.

The test subjects' deliveries were recorded with Audacity, while an SMI RED250M eye-tracker was used also to record pupillometry data. A LOG file, automatically created by InterpretBank, served as a reference to check what terms had been looked up by the test subjects. The same was done manually for the trials in which Word and Excel were adopted, using the Gaze Replay recordings. The interpretations were then transcribed using Partitur Editor, the transcription tool of the Exmaralda suite, and then analyzed. Before presenting the method used for this analysis and its results in section 4, we will describe the main features of the speeches used for data collection, with a focus on stimuli distribution and morphological complexity.

Which  
section?

## 4 Stimuli features and distribution

While asking the test subjects to interpret single terms would have eliminated the time constraint typical of simultaneous interpreting, working with authentic, unedited speeches would have introduced too many variables in our experiment. For this reason, we decided to adapt Seeber2011a's methodology (Seeber2011a) by creating ad-hoc speeches made up of sentence clusters. This method presents three main advantages. First, it enables us to focus our investigation on the target sentences (i.e. the ones which include the stimulus). Second, it makes it easier to work with comparable speeches, as they have the same structure. Third, it gives the test subjects the impression that they are interpreting a speech, rather than disconnected sentences, thus helping us retain a certain degree of ecological validity. Each sentence cluster is composed of a general, introductory sentence, followed by the target sentence containing the stimulus, followed by a third sentence which, like the first one, does not contain specialized terminology. The structure is repeated throughout the speech, so that each stimulus is separated from the next one by two sentences. Here is an example from speech n.1:

- (1) So we need to change this basic trend and this is why the urgency is there.

In our policies, we should definitely address the need to improve **vehicle efficiency**.

But there is still much more we can do, in many other areas, as you are aware.

At the EU level, there is another policy option that can help us.

By focusing, for instance, on **woody biomass fuels**, we can truly make a difference.

They have the potential to help us respond to the challenges we're facing.

Each speech prepared for data collection contained 36 terms, 12 of which are unigrams (e.g. "bioenergy"), 12 bigrams (e.g. "energy poverty") and 12 trigrams ("pressurized water reactor"). This variable was introduced because the length of the stimuli is expected to also play a role in the usability of the tools. We expect there to be differences between tools when a more morphologically complex term is looked up – it should be more difficult to find a trigram when using a Word or an Excel glossary than when working with a CAI tool.

Of each group of stimuli, 6 are placed at the end of the sentence and 6 in the middle of the sentence. This was done to verify whether the stimulus position has an impact on cognitive load and on the test subjects' behavior in querying the glossary. We expect the stimuli placed at the end of the sentence to lead to a lower increase in cognitive load and to be looked up more often, thanks to anticipation.

Of the 6 terms placed at the end of the sentence, 3 should require a query in the glossary, because they are less frequent and thus probably unknown to the participants, and 3 should not.<sup>10</sup> The same is true for the terms placed in the middle of the sentence. Half of the stimuli in each speech should therefore require a query and half should not. This variable was introduced because we want to verify whether CAI tools, which are usually deemed to be user-friendlier and to take up fewer cognitive resources, allow participants to perform more queries without leading to a higher number of errors or omissions. Figure 8 sums up the features of the stimuli and their distribution in each speech. For each speech, the stimuli can thus be classified according to their features, for future analysis. Table 1 shows an example of this classification for the stimuli in speech nr. 1.

## 5 Stimuli validation

One of the aims of the exploratory study was to verify whether the stimuli prepared for data collection, to be used also in a future experiment involving a larger sample, elicited the reaction we expected from the test subjects, i.e. a query in

---

<sup>10</sup>This classification was based on the frequency of the terms as per the 2015 news corpus, the 2012 web corpus (UK) and the 2016 Wikipedia corpus for the English language (Projekt Deutscher Wortschatz, <http://wortschatz.uni-leipzig.de>).

Table 1: stimuli classification – speech 1

Stimulus	Position	Morphological complexity	Glossary search needed (GS)
bioenergy	E	1	
security of supply	M	3	
gasoline	M	1	
conventional fossil fuels	M	3	
vehicle efficiency	E	2	X
woody biomass fuels	M	3	X
liquid biofuels	E	2	
rapeseed methyl ester	E	3	X
transesterification	E	1	X
short-rotation coppice	E	3	X
black liquor	E	2	X
corn stover	E	2	X
lignocellulosic solid biomass	E	3	X
gasification	M	1	X
gasifier	E	1	
green charcoal	M	2	X
briquettes	M	1	X
biofuels sector	M	2	
soil protection	E	2	
petroleum	M	1	
greenhouse gas emissions	E	3	
EU biofuels directive	M	3	X
indicative targets	M	2	X
incentives	E	1	
set-aside land	M	3	X
arable land	M	2	
solar power	M	2	
second-generation biofuels	E	3	
switchgrass	E	1	X
first-generation biofuels	E	3	
residue cake	M	2	X
milling	E	1	X
malting	M	1	X
overall energy demand	M	3	
renewables	M	1	
energy mix	E	2	



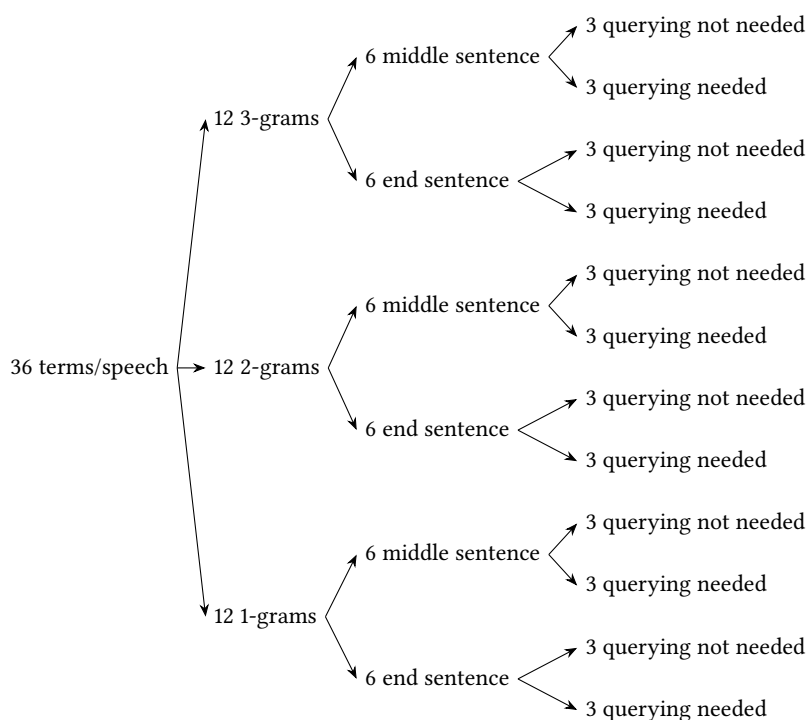


Figure 8: Features and distribution of the stimuli used for data collection

the glossary. This is necessary to make sure that enough queries are performed in the glossary to provide sufficient data for a comparison between the three terminology management solutions we focus our analysis on – Word glossaries, Excel glossaries and second-generation CAI tools. While a certain degree of inter-subject variability can be expected, we must verify whether our a-priori classification of the stimuli holds true on a general level. This is the focus of the first part of the transcription analysis, which will be presented in section 4.1.

Another goal of our research is to verify whether the use of a CAI tool leads to better terminological quality in comparison to more traditional terminology management solutions, e.g. Word and Excel glossaries. First observations made in our sample are briefly discussed in 4.2, where we also provide a framework to analyze errors and omissions in relation to the tools used for glossary query.

§??presents the results of our observations in relation to the strategies adopted by the test subjects to interpret the stimuli. Given the small size of the sample,

Do you mean 5.1?

This section

You linked to §4.3 here, but it doesn't exist. Maybe you

with this exploratory study we aim at developing a methodology to be used for further research, rather than at drawing conclusions, which will require a larger set of data.

## 5.1 Stimuli classification

As previously stated, half of the stimuli were classified as needing a glossary query. In order to verify whether this was true, the sample was checked for the total number of terms searched, the number of terms searched that were classified as needing to be searched in the glossary (“QN”) and the number of terms searched that we did not expect to require a query in the glossary (“NO QN”).

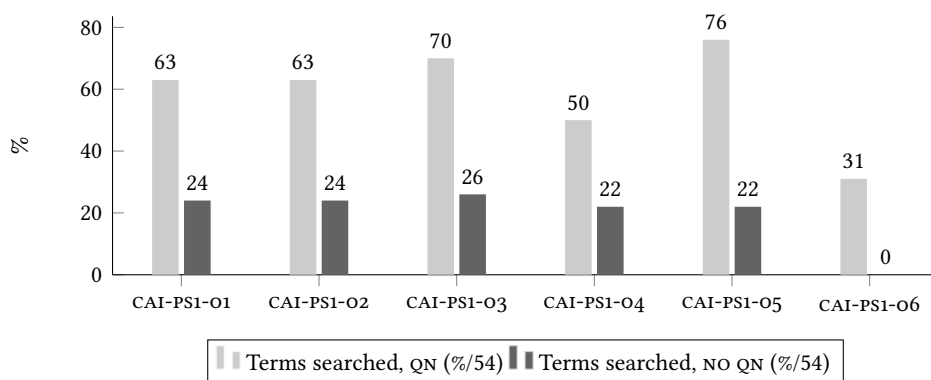


Figure 9: Search behavior per stimuli category

As shown in Figure 9, the percentage of terms classified as needing a query that were actually searched varies among the test subjects, while it is quite similar in the case of terms classified as not needing a query. A notable exception is test subject CAI-PS1-O6, who searched a much lower number of terms than the other test subjects. The percentages are very similar for the German natives (participants CAI-PS1-O1, CAI-PS1-O2 and CAI-PS1-O3), although they looked up different terms.

We also verified which terms classified as needing a glossary query had not been looked up by any subject. 5 terms out of 54 were not looked up by anyone and should therefore either be moved to the non-query category or replaced by more specialized, less frequent terms. Of the terms classified as not needing a query in the glossary, only 1 out of 54 was looked up by all test subjects. It should therefore either be classified differently or replaced.

If we take into consideration the position of the stimuli in the target sentences,

something interesting emerges from initial data analysis, which deserves further exploration in a bigger sample, especially in correlation with pupillometry measures. While the difference is, for some test subjects, more evident than for others, the stimuli placed at the end of the sentence seem to elicit more queries than the stimuli placed in the middle of the sentence (see Figure 10). This might be explained with the fact that, when a term is placed at the end of the sentence, anticipation might lead the participants to prepare themselves to adopt a coping mechanism, such as a glossary query. The “preparation” could also result in a sentence structure that favors a glossary query, requiring less restructuring or making it difficult to omit the term completely. This could, however, result in higher cognitive load, because if the query is not successful, more cognitive resources will be needed to adopt a different strategy, possibly affecting the rendition of the following sentences. A stimulus placed in the middle of the sentence could prompt the interpreter to immediately choose a strategy different than consulting the glossary available, such as generalization or the use of a synonym. While this may lead to a less precise rendition of the original, it may also come with lower cognitive load experienced.

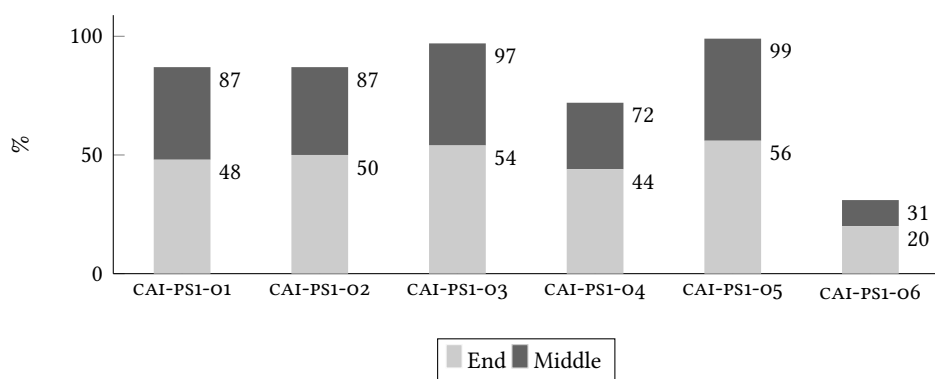


Figure 10: Stimulus position and percentage of terms searched. Percentage expressed on a total of 54 terms per category.

Even though the stimuli classified as requiring a query were equally distributed in terms of position – half of them placed in the middle and half at the end of the sentence – the difference in the search behavior might also be due to the terms themselves, rather than only to their position. This can be further tested on a larger sample, by switching the position of the stimuli or by using a different set of stimuli.

If we take into consideration the morphological complexity – here defined as

the number of elements making up the terms<sup>11</sup> – we notice that unigrams are searched more often than bigrams and trigrams in the case of the stimuli classified as not needing a query (see Figure 11). This might be explained with the fact that, when faced with a bigram or a trigram, participants need to decide which element of the term should be looked up, which requires additional cognitive resources. For this reason, they might choose to directly adopt a different strategy. A unigram does not require them to make this decision, and so the act of querying the glossary is more straightforward. No clear trend can be identified for the stimuli that should require a query.

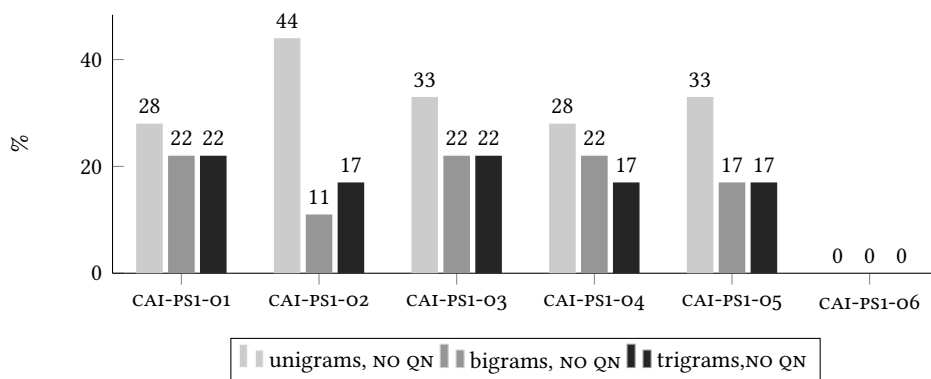


Figure 11: Morphological complexity and percentage of terms searched (terms not needing a query). N=18/category

All in all, our a-priori classification of the stimuli was confirmed by the sample, if we exclude the outlier CAI-PS1-O6. Further research will be needed to check our hypotheses on the role played by the position and the morphological complexity of the stimuli.

## 5.2 Tools used and precision of renditions

With the aim of gaining initial data on how the tool used influences the precision of the test subjects' renditions, we compared the level of precision observed for the Word glossary, the Excel glossary and InterpretBank, when a glossary query was chosen as the strategy to interpret the terms. Our classification of the renditions is loosely based on Wadensjö1998 and is made up of 3 main categories:

**Close renditions (precision 2 – P2):** no information lost, precise rendition, use of equivalent as per glossary or adequate synonym

<sup>11</sup>For instance, a trigram is considered morphologically more complex than a bigram.

**Acceptable renditions (precision 1 – P1):** some information is lost (e.g. through paraphrasing, the loss of an adjective in trigrams, a drop in register), but the general meaning is maintained

**Zero/unacceptable rendition (precision 0 – P0):** the rendition completely or largely deviates from the original message (the content is different), or the information is not present (zero rendition).

This classification certainly presents some degree of subjectivity, but it is nonetheless useful as a broad guideline to evaluate the precision of the test subjects' deliveries.

Figures 12, 13 and 14 sum up the degree of terminological precision achieved when performing a glossary query with Word, Excel and InterpretBank glossaries.

Inter-subject variability is too high to draw initial conclusions on this aspect, but Excel seems to lead to the worst performance, since we can notice more occurrences of zero renditions or unacceptable renditions than for Word and InterpretBank. This is probably due to the fact that, when working with Excel, our test subjects did not have the possibility to view all the results of a query, but only to manually skip to the next occurrence, which might make the query too cumbersome to be performed in the very small amount of time available to the interpreter. InterpretBank seems to perform slightly better than Word, but this should be further verified. Usability probably plays a role in this respect, so eye-tracking measures will be key in determining how the user interface facilitates or hinders the identification of the equivalent needed.

As for the morphological complexity, we expected queries performed with InterpretBank to be more effective – leading to a higher level of precision – than queries performed with Word and Excel, especially for more complex terms (trigrams). In the small sample analyzed in the pilot study, queries performed with InterpretBank lead to higher precision for unigrams in 5 cases out of 6. The only exception is participant CAI-PS1-06, for whom we have very few data points when compared to the rest of the sample. For bigrams and trigrams the results are less uniform – queries with InterpretBank are more effective than Word and Excel in 3 participants out of 6. While the sample analyzed is too small to draw conclusions, this aspect can be further analyzed in a larger sample, where differences might be significant.

It should be noted that in order to facilitate the analysis, we first took into consideration only the terms and the content conveyed by them, not by the whole sentences. This analysis hence remains focused at a microscopic level, that of

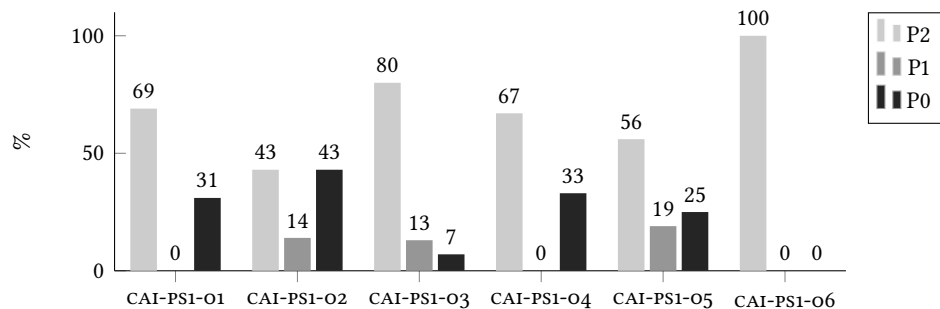


Figure 12: Precision of renditions with Word

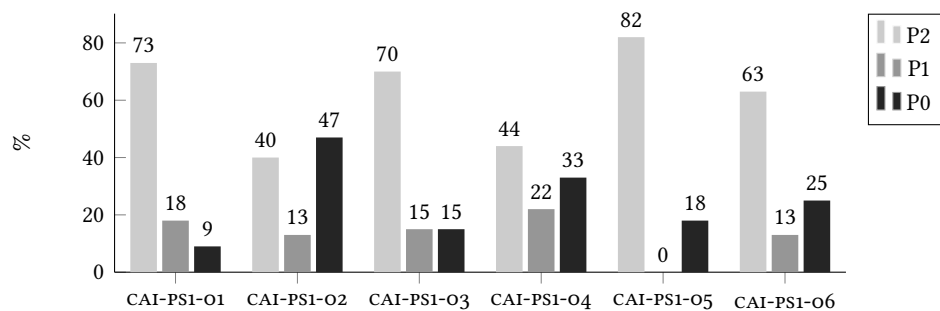


Figure 13: Precision of renditions with Excel

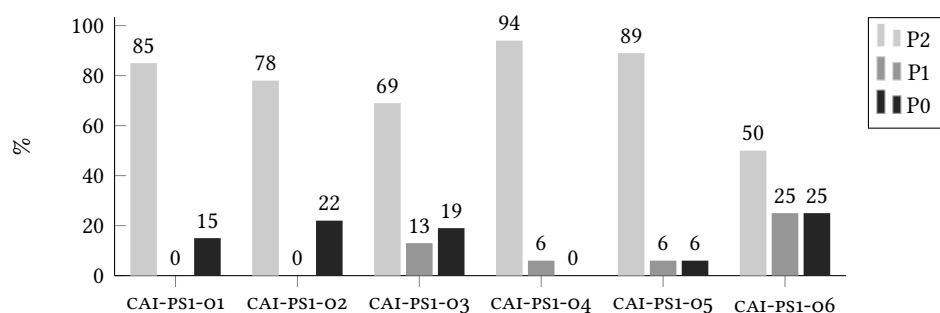


Figure 14: Precision of renditions with InterpretBank

terminology. Since the ultimate goal of CAI tools is that of improving the overall quality of the rendition, we deem necessary to expand our analysis to the sentence level, to verify whether a higher level of precision achieved in the rendition of the terminological unit results in a correct and complete rendition of the sentence it is embedded in – and of the following ones – or whether, on the contrary, the query, despite being successful, leads to errors or omissions. To this aim, the transcriptions of the test subjects' renditions were annotated following Barik1971's (Barik1971) classification of omissions, additions and errors in *si*. In the data analysis, we decided to focus on three categories, which represent the most serious issues encountered in the rendition, namely E4 (substantial phrasing change), E5 (gross phrasing change) and a third category which corresponds to a complete omission of the sentence, which we labeled as M5.<sup>12</sup> On the other hand, to make our analysis easier, we grouped in one category the renditions that did not present any issue or only presented less serious issues, such as skipping omissions and mild phrasing changes. The classification of errors and omissions provides an element of subjectivity, which might be constrained by taking into consideration only clearly wrong sentences or total omissions.

Given the small sample of the study and the subsequent high level of inter-subject variability in terms of amount of terms searched, we were not able to identify any clear trends from this data alone. The statistical significance of the data will have to be verified on a bigger sample. Nonetheless, the pilot study was useful to define a working method that can be applied to further research and possibly expanded to also take into consideration the features of the stimuli.

### 5.3 Tools used and interpreting strategies

We conclude our analysis by looking at the strategies adopted, to establish whether a correlation can be found with the tools used. The classification of the interpreting strategies is based on Bartłomiejczyk2006. In our analysis, we focused on the "strategies of production" (ibid.), which can be observed by analyzing the product of *si*, while we did not take into consideration overall strategies, which would require additional methods to be identified. From the analysis of the transcriptions, 10 main strategies, or coping-tactics (Gile1995), emerged:

1. Glossary search (GS)
2. Approximation (A): use of a synonym or a closely related term

---

<sup>12</sup>See Barik1971 for a complete classification of errors, omissions and additions in *si*.

3. Compression (C): use of a hyperonym, some precision is missing
4. Omission (o): not strictly considered a strategy, it is mostly unintentional
5. Paraphrase (P)
6. Reproduction (R): no translation, the term is reported as in the source language
7. Transfer (T): ad-hoc translation
8. Syntactic transformation (ST)
9. World knowledge (WK): reference to one's pre-existing knowledge
10. Substitution (S): the term is replaced by another term, not related to it.

Figure 15 reports an overview of the strategies used by the test subjects for all tools and all stimuli. Data clearly show that, with one exception, a glossary query was the strategy most used by the test subjects. This can be easily explained by the fact that the test subjects had not prepared for the assignment. The second most used strategies are approximation, omission and world knowledge.

The third most used strategies are world knowledge, paraphrase and omission.

If we look at the strategies adopted when using different tools to look up terms in the glossary, we can notice that, when using InterpretBank (see Figure 16), a glossary query is the favorite strategy, except for one subject (the same as in the general analysis), who seems to resort mainly to approximation. The second most used strategies are omission and approximation, while the third most used strategies are world knowledge, approximation and omission.

Querying the glossary was the favorite strategy also when Excel (Figure 18) was used, in 4 cases out of 6, while the other two resorted, respectively, mainly to omission and paraphrase, and to world knowledge. There is not a clear preference as to the second most used strategy, while the third most used, in 4 cases out of 6, is world knowledge, followed by approximation and both omission and compression.

In the third case, in which the test subjects could look up terms in a Word glossary (see Figure 17), a glossary query also seems to be the favorite strategy, while paraphrasing is the second most used strategy in the sample. The third most used strategy is omission.

Even though a glossary search was the preferred strategy by almost all participants irrespective of the tool used, the percentage of queries performed with



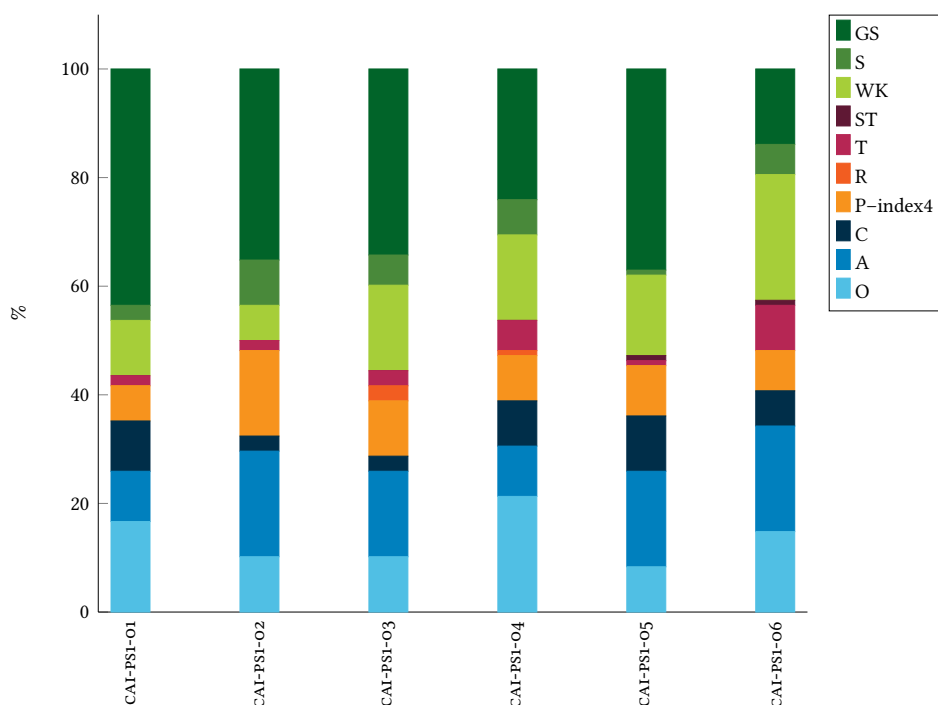


Figure 15: Overview of strategies used (all tools and stimuli)

Todo for Langsci: fit to typeblock

InterpretBank seems to be higher across the board, except for one participant. By looking at these initial data, we can thus hypothesize that test subjects find it easier to perform a glossary query when using InterpretBank, probably due to better usability, and that reference to previous knowledge, approximation, paraphrasing or outright omission are the preferred coping tactics when the glossary is not queried. This should be tested on a larger sample by triangulating data from transcriptions with eye-tracking data.

## 6 Conclusions and further research

The paper presented first results from an exploratory study aimed at developing a research methodology to investigate the use of computer-assisted interpreting tools during simultaneous interpreting. The pilot study is part of a PhD research project that aims at collecting data both on the procedural and the terminological

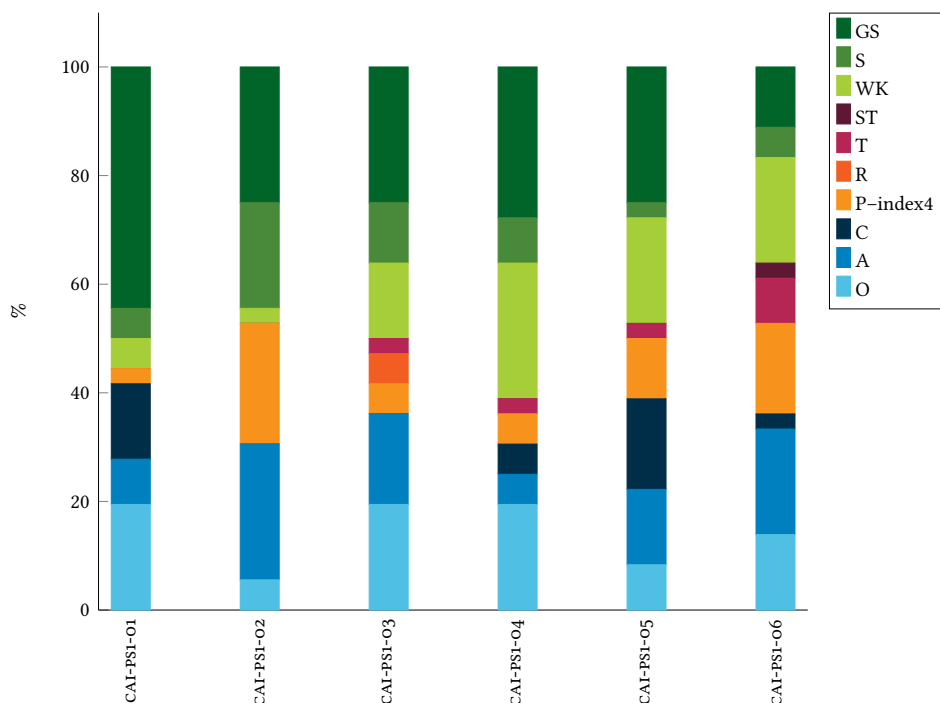


Figure 16: Strategies adopted – InterpretBank glossary

aspect of SI with CAI, combining product- and process-based measures.

After discussing the theoretical framework chosen for the study, we presented our main hypotheses on cognitive load during SI with CAI. In particular, we expect cognitive load to be higher during SI with CAI than during traditional SI, but to be lower for CAI tools such as InterpretBank than for traditional terminology management solutions like Word and Excel glossaries. We also expect the terminological quality to be better when a CAI tool is used. While the hypotheses on cognitive load will require the analysis of pupillometry and eye-tracking measures to be verified, the analysis of the interpretations helped shed some light on the terminological quality of SI performed with the support of CAI tools and of traditional table glossaries.

First data from the transcriptions of the test subjects' deliveries have proved helpful to verify the adequacy of the stimuli created for the experiment, showing that the a-priori classification of the stimuli used is overall confirmed by the test subjects' search behavior, in particular when it comes to the stimuli classification into terms expected to require a glossary query and terms not requiring a query.

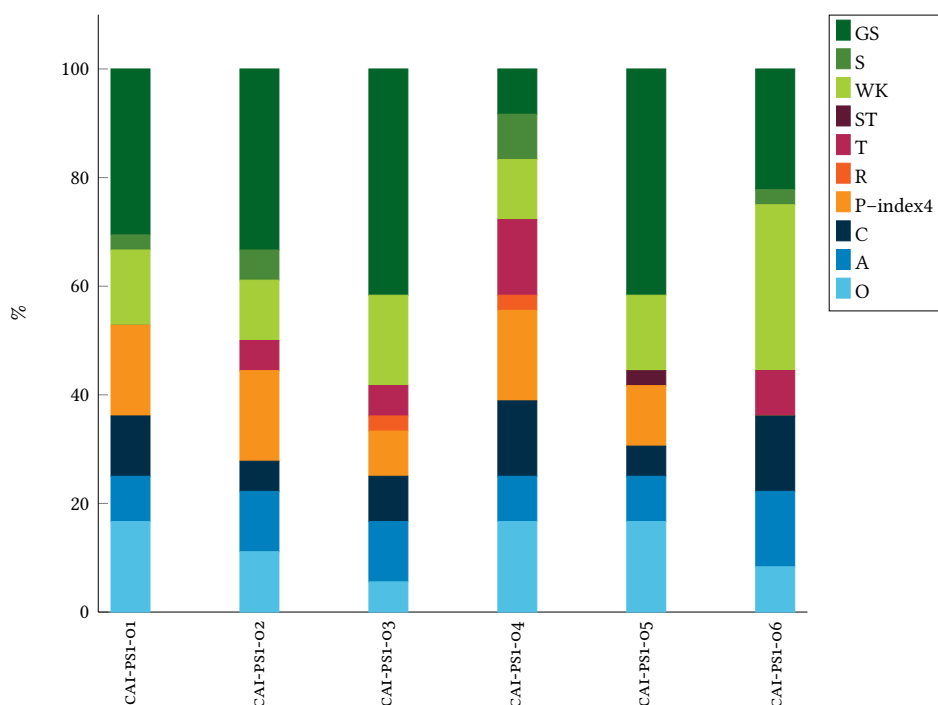


Figure 17: Strategies adopted – Word glossary

The position of the stimuli seems to play a role in the search behavior, while their morphological complexity does not seem to have a significant impact on it. InterpretBank seems to provide the highest degree of precision, and the glossary query appears to be the favorite kind of strategy to apply to cope with specialized terms when InterpretBank can be used to search for terminology. All of these aspects will need to be further investigated in future studies.

Further analysis of process-related and usability data, in particular of pupillometry in triangulation with eye-tracking measures, will be necessary to gain information that can shed some light on the hypotheses on cognitive load and help formulate further hypotheses.

Finally, future studies should also include the option to query the glossary through automatic speech recognition, which can be expected to be the most “cost-effective” option in terms of cognitive load added and level of precision, coherence and cohesion achieved in the interpretation.

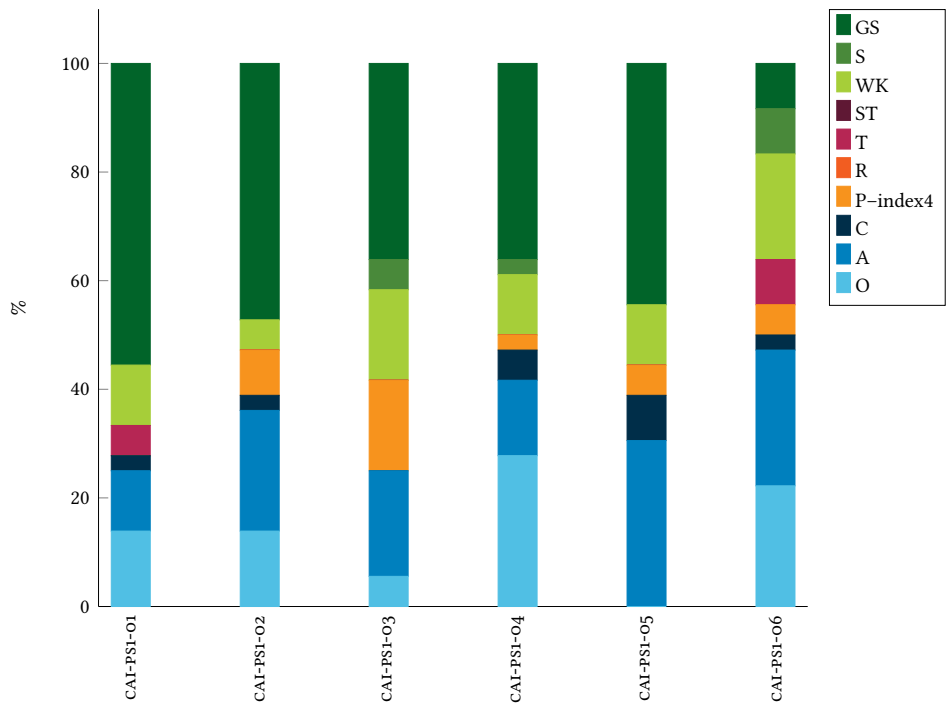


Figure 18: Strategies adopted – Excel glossary

## Chapter 4

# Experimenting with computer-assisted interpreter training tools for the development of self-assessment skills: National Parliament of RSA

Elizabeth Deysel

National Parliament of South Africa

Harold Lesch

University of Stellenbosch

This article explores the use of CAIT as a tool in the development of self-assessment skills in interpreting performance. The aim of this pilot study is to investigate and evaluate the effectiveness of the CAIT software in the development of self-assessment skills of practicing interpreters in the National Parliament of the Republic of South Africa. The research design for this article comprise an evaluation study approach, based on an experimental intervention design, i.e. the intervention by exposure to CAIT for purposes of self-assessment. In order to collect data that would address the research questions, a questionnaire, an experiment and interviews were used. The experimental group was exposed to the software, Black Box, in order to measure the impact thereof on the development of their self-assessment skills. The results indicate that the experimental group practicing interpreters which were exposed to the software, displayed an improvement in their self-assessment skills and they indicated a better understanding of the criteria which are important in the assessment of interpreting performance as well as a better awareness of the strengths and weaknesses in the interpreters' interpreting performance. The study concludes that CAIT may prove a viable tool also for in-house training and development of self-assessment skills of professional interpreters.

Please shorten the abstract such that the main text begins on the title page



## 1 Introduction

Computer-assisted interpreter training (CAIT), as a relatively new field in interpreting studies, explores the implementation of information and communication technologies (ICT) in the training of interpreters. Currently very little, if any research has been conducted on CAIT within the South African context. International research on CAIT and its application in the development of self-assessment skills has focused mainly on its implementation within institutions of higher learning as a tool in the training of student interpreters. There has been no focus on the possible use in the training and self-assessment of the practicing interpreter. These CAIT tools may also prove useful when utilized for self-assessment skills development within institutions that employ interpreters on a permanent basis.

The curriculum for training interpreters has seen a significant evolution over the past two decades. The implementation of information and communication technologies (ICT) in interpreter training is a useful additional tool in the interpreting curriculum. ICTs provide a variety of tools that can enhance the teaching and learning of interpreting and how trainers go about the process of training potential interpreters. This contention is borne out by the number of scholars who have recently shown an interest in and published texts on the subject. In this regard, the contributions of Lim2014, Pinazo2008, Gorm2007, Sandrelli2007b, Lee2005 and Sandrelli2015; Sandrelli2007b; Sandrelli2002 are relevant. The aforementioned studies led to insights that the implementation of computer-assisted interpreter training (CAIT) in the training of interpreters may be desirable and an appropriate addition to traditional training methods as it holds a number of advantages for both the trainee and the trainer. One of the main advantages highlighted in these studies is the shift towards and emphasis on learner autonomy.

The aforementioned studies were conducted within the context of implementation in the interpreting curriculum and the training of student interpreters at institutions of higher learning. However, these tools may also prove useful when utilized by freelance professional interpreters and within institutions that employ professional interpreters on a permanent basis. This study poses the question whether these CAIT tools are effective in the development of self-assessment skills in the professional interpreter. This question was approached by utilizing the CAIT software, Black Box<sup>1</sup>, within a professional interpreting environment

---

<sup>1</sup>In 2002, Melissi Multimedia Ltd (UK) collaborated with the University of Hull (UK) on the design of a digital language laboratory. As part of this development, a dedicated interpreter training module, called *Black Box*, was included. The software *Black Box* was developed as a

chosen as the Interpreting Unit of the National Parliament of South Africa. The effectiveness of this training software as a self-assessment skills development tool for practicing interpreters was evaluated.

In the context of the latter area of interest, this article presents research conducted on the utilization of CAIT as a tool for the development of self-assessment skills in professional interpreters. The article is organized as follows: firstly, the rationale concerning self-assessment and computer-assisted interpreter training is highlighted; secondly, an overview of the methodology is provided thereafter the results in pertaining to the hypotheses are discussed.

## 2 Theoretical background: Self-assessment and CAIT

In this section, information is provided on self-assessment in interpreting, computer-assisted interpreter training and background on the professional interpreter.

### 2.1 Self-assessment in interpreting

Regehr1996 define self-assessment as the ability of each individual to identify his or her own relative strengths and weaknesses. They also offered a reconceptualization of self-assessment that shifted from a focus on the individual's ability to rate themselves relative to their peers and moved on to explore the ability of the individual to identify their own strengths and weaknesses relative to each other. It is suggested that the ability to identify areas of performance that require the greatest degree of improvement would lend greater efficiency to self-directed learning efforts.

Riccardi2002 states that the training period is of key importance for introducing future interpreters to the habits of recognizing their strengths and weaknesses. Interpreter training courses are intensive in nature and training is complimented by additional self-study hours. However, self-study hours as in the case of experiential learning bear the risk of being of little use if there is no reflection upon the experience. Sandrelli2007a states in this regards that "if unsupervised practice sessions are to be useful, students need to be able to assess their own performance and identify their weaknesses. Indeed, the development of self-assessment skills is an essential component of interpreter training".

There is agreement in the research by Pinazo2008 when contending that the training period is vital for introducing interpreters to self-assessment skills and

---

commercial product by Melissi Ltd. in 2005.

that the integration of self-assessment skills will also have positive effects on learners' attitudes to self-criticism and performance.

Fowler2007 emphasizes the importance of self-assessment skills in interpreting when she explains that after training most interpreters remain isolated throughout their professional lives and the process of monitoring of the interpreter is likely to be left to the interpreter themselves. If the interpreter is not self-aware, and has neither skill to be able to assess or evaluate their own performance nor take action to improve upon weaknesses, the service user will suffer the consequences. She elaborates that self-assessment in interpreter training therefore fosters good professional habits in the interpreter. This is also mentioned by Lee2005 when he states that "self-assessment is not only important during the training phase of interpretation, but it is critical to professional interpreters as well". He further explains that "freelance interpreters are often left to check their own interpretation quality and find measures for improvement" (Lee2005). The research from Sandrelli2007a has also highlighted that "self-assessment skills and the ability to assess other interpreters' performances are essential for trainees both to ensure progress and to maintain quality standards in their future careers as professional interpreters".

The research conducted on the subject (Riccardi2002; Lee2005; Sandrelli2007a; Fowler2007; Pinazo2008) indicate that the development of self-assessment skills, is essential in interpreter training. It is concluded that the development of self-assessment skills in the student interpreter will allow for the ability of the individual to recognise his or her strengths and weaknesses and apply appropriate coping mechanisms to enhance the parts of their performance that need improvement. The development of these self-assessment skills will foster good professional habits which can be used to monitor their progress and ensure quality standards in the future career of the professional interpreter.

## 2.2 Computer-assisted interpreter training

Sandrelli2007a indicate that since the 1990s several independent projects were undertaken that shaped the gradual development of what has come to be known as CAIT. This development has resulted in the division of CAIT into what is known as *integrative CAIT* and *intelligent CAIT*. *Integrative CAIT* entails the implementation of ICT in interpreter training focused on the creation of digital speech repositories in the form of databases such as the *Interpreters' Information System* (IRIS) developed at the University of Trieste in mid 1990s (Carabelli1997) and *Marius* developed at the University of Granada in 2001 (Pöchhacker1994). These projects collected digital training materials and streamlined these resources for use by stu-



dents in self-study sessions. They have been labelled as *integrative CAIT* – as these projects “exploits the integration of audio, video and textual resources to provide students with suitable material for classroom use of self-study” (Sandrelli2007a). On the other hand, *intelligent CAIT* involved the development of authoring programs such as *Interpretations* and *Black Box*, which enables interpreter trainers to create various types of exercises.

**Berber2010** was one of the first who investigated the use of Information and Communication Technologies in interpreter training and elaborates on the use of these tools as means<sup>2</sup> or pedagogical tools. Even though ICT does not facilitate interpreting immediately but enhance learning over time. She also integrates the Effort Model (Gile1995) and which of the efforts can be backed up by ICT. **Berber2010** concludes that ICTs in general support the efforts presented in the Effort Model and that information technology in the form of interpreter training tools are specifically aimed at the second effort (production) of Gile’s Effort Model, where the student can “listen to him/herself repeatedly for self-evaluation and improvement of production skills”. In her research, **Berber2010** indicated that the types of ICT which are being used for self-training are mainly traditional: booths, language labs, digital recordings, video and audio recordings, internet, PCs, e-learning platforms.<sup>3</sup>

CAIT tools include CR-ROMs, speech repositories, speech and recording databases and authoring tools such as the software program *Black Box*. The aforementioned software program allows interpreter trainers to create and develop their own set of interpreting exercises for use by individuals and interpreting students in their own time for their self-study sessions. Research that has been conducted on the topic of CAIT (Sandrelli2007a; Pinazo2008; Lim2014) indicates that implementing CAIT in the training of interpreters not only enhances the teaching and learning of interpreting, but also enables the creation of a realistic practice environment in which student interpreters are able to develop their self-assessment skills by listening to their own interpreting and reflect upon it. **Bartlomiejczyk2007** indicates that self-evaluation by means of critically listening to one’s own recorded interpreting has often been suggested as a useful method of quality control. The development of self-assessment skills enables the student interpreter to identify strengths and weaknesses, apply appropriate coping strategies and monitor their progress and performance.

---

<sup>2</sup>The term “means” indicates that the ICT tools are used to practice and develop skills – as opposed to being used for support during or in preparation of actual interpreting.

<sup>3</sup>Specific brands of equipment are X-class, Melissi Blackbox, Sanako, Dialang language tests, DEYA lab, Trados, Audacity, BNc online and Brähler.

In her research, Sandrelli2007b discusses the development of the interpreter training prototype, *Interpretations*, and how that prototype was improved to become the CAIT authoring tool known as *Black Box*. In 2002, Melissi Multimedia Ltd (UK) collaborated with the University of Hull (UK) on the design of a digital language laboratory. As part of this development, a dedicated interpreter training module, called *Black Box*, was included. After interest was shown by interpreter training institutions, Melissi Multimedia Ltd decided to develop *Black Box* as a stand-alone program, and it was released in March 2005. *Black Box* is an authoring program – this means that the interpreter trainer has complete control over the resources contained within the program. The software was developed with a hierarchy of how materials are structured. There may be different courses and each of these courses may contain different modules which will then each contain different exercises. The software’s authoring function allows the interpreter trainer to create these different courses, modules or exercises which may comprise simultaneous, consecutive (including liaison interpreting) as well as exercises for sight translation.

The different exercises suggested by the developers are:

- a. shadowing and closing;
- b. paraphrasing;
- c. sight translation;
- d. simultaneous interpreting;
- e. simultaneous interpreting with text and
- f. consecutive interpreting.

Potentially it allows one to compile exercises the way you want them to be, by combining text, video and audio. These are suggested activities that take into account an interpreter’s learning path in a specific course. Sandrelli2007a also indicated that the *Wizard* makes it possible to add many more resources, including instructions to students, a written translation of the speech, written exercises (comprehension questions, text analysis exercises) and a teacher’s interpreted version of the speech. Teachers can also manipulate the sound stream by adding an echo effect or sound distortion in order to simulate realistic working conditions. The source text transcripts can be annotated by adding a hot footnote. Students read the note made by the teacher simply by moving the mouse over the word. In the sight translation exercises the text is presented to students in a scrolling cylinder which advances at the pace established by the teacher.

converted  
from  
inline  
enumer-  
ation to  
list

## 2.3 The professional interpreter?

Since this research study focused on the utilization of CAIT beyond institutions of higher learning in to interpreting practice, the term “professional interpreter” was often referred to. It was thus deemed necessary to provide a definition on the concept “professional interpreter”. Using time as a measure to achieve professional status, in an article by Sandrelli2015, reference is made to Moser-Mercer (in Motta2006) who estimates that 3000–5000 hours of deliberate practice are required in order to achieve professional levels of expertise in interpreting. The footnote of the mentioned article indicates that AIIC (International Association of Conference Interpreters), admits new members with a minimum of 150 days of work experience.

In her article on *Language practitioners and Standards*, Feinauer2005 states that the characteristics of a profession are “mastery of a particular skill through education and training, acceptance of duties to a broader society than merely one’s clients/employers, objectivity and high standards of conduct and performance”. She goes further and defines the profile of a professional as an individual “trained to recognise standards of competence, adheres to a recognised code of practice and enjoys the support and regulation of a professional structure” all the while stating that professionalism is a relative term.

In summary, the term “professional interpreter” is therefore defined as an interpreter presumed to not simply be competent but having mastered their skill with prior experience and/or training in interpreting and adhering to high standards of conduct supported by a code of practice.

## 3 Methodology

This section provides information regarding the research design, respondents and the methods (questionnaires, experiment and interviews) used to collect the empirical data.

### 3.1 Research design

Using the above background as the point of departure, the primary objective of the research was to investigate and evaluate the effectiveness of the CAIT software, *Black Box*, in the development of self-assessment skills of professional interpreters in the National Parliament<sup>4</sup> of South Africa. To address the primary

---

<sup>4</sup>The National Parliament of South Africa makes use of interpreting into the eleven official languages during their sittings as well as Sign Language. The eleven official languages are

research objective as stated, the following secondary research questions were explored:

- To what extent does training in self-assessment for interpreters give a better understanding of the strengths and weaknesses of their interpreting?
- To what extent does training in self-assessment for interpreters give a better understanding of the criteria used in the evaluation of interpreting performance?
- What is the correlation of the self-assessment ratings between the experimental group and the ratings from the expert assessor post-experiment?
- What is the correlation of the self-assessment ratings between the control group when compared to the experimental group post-experiment?

The research design most suitable for this study comprised an evaluation study approach, based on an experimental intervention design, i.e. a type of study in which participants are assigned to groups that receive one or other intervention or no intervention so that the effects can be evaluated of the intervention. An intervention research includes studies in which researches follow a systematic change in the condition to determine the effects on a physical capacity, skill or performance. In the evaluation of the effectiveness of the CAIT software, *Black Box*, the software was utilized as an intervention in the form of technological innovation in voluntary in-house training to support professional interpreters in their professional development. It should be noted here that the in-house training formed part of this research study and was not initiated or permanently implemented by the Parliament of South Africa. Therefore, the researchers' personal PC was used in the sessions which has one licensed copy of *Black Box*. The participants were exposed to self-assessment sessions on *Black Box* in individual sessions where they received the same brief and instructions beforehand. The sessions were conducted during lunch hours in a sound-proof room with two sound-proof doors.

The empirical study sought to obtain quantitative and qualitative data. This meant that the core method was of a quantitative measure, while the supplementary method was of a qualitative nature and was used to extend the findings of the quantitative data. The quantitative data was collected from the experiment,

---

Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga.

which required the interpreters to complete self-assessment grids (see Addendum 5) in both the control and experimental groups. An investigation by means of a questionnaire (see Addendum 5) and interviews (see Addendum 5) formed part of the qualitative follow-up to investigate the outcomes from the quantitative data.

Langsci:  
Check  
link to  
addendum

It should be emphasized that this is a pilot study as the sample size of respondents is extremely small and may contribute to the data not being statistically significant. To put this into context the following background information should be noted.

According to Human Resources of National Parliament of South Africa (Moorad2017) 38 language practitioners were employed within the Interpreting Unit at the time of conducting this study. Of these 38 practitioners, three were Sign Language interpreters. These interpreters could not participate in the study, as the software, *Black Box*, does not make provision for video recording. This left 35 language practitioners available for participation in the study.

The institutional permission the researcher received from Parliament to conduct the research within the Interpreting Unit stipulated that data may only be collected outside of work hours. The researcher agreed to this stipulation, which meant the lunch hour was used for data collection. The experimental part of the study – that involved the self-training sessions on *Black Box* – would take up to 30 minutes per person per session. With the time allocation for the experiment in mind, the researcher calculated that only five respondents per week could form part of the experiment. A limitation resulting from this agreement is that the researcher observed that collecting data from participants outside of working hours i.e. during their lunch breaks, may discourage some respondents from participating in the study and that reluctance may result in the entire population in the unit of analysis not participating in the data collection.

When surveying only a sample of the population, researchers have to consider margins of error and confidence levels of the data that is collected: the margin of error is the amount of error which can be tolerated, while the confidence level is the amount of uncertainty that can be tolerated. The margin of error for this study was set at 25% while the confidence level in the study was set at 90%. Given the population of 35 possible respondents, the sample size was calculated at nine.

For this study, it was decided that anything above 26% as a margin of error would be too high. A margin of error of 25.91% would mean that eight respondents would form part of the study. The researcher had to bear the possibility of discouragement of some respondents in mind, and thus decided to send the questionnaire to double the amount, resulting in 16 interpreters receiving the link to

the questionnaire. Only ten of the 16 respondents had completed the questionnaire.

The experimental method in this study involved:

- Selecting a group of respondents<sup>5</sup> who fit into the category of ‘professional interpreter’
- Dividing them into an experimental group and a control group using the quota matching method
- Exposing the experimental group to a stimulus – in this case four self-assessment sessions on *Black Box*, and lastly
- Observing and measuring the effect of the stimulus on the respondents. The experiment itself entailed pre- and post-testing of the respondents. The pre-testing tested the respondents to determine their self-assessment skills.

The experimental group was then exposed to self-assessment sessions which served as the intervention. Finally, post-testing was conducted to determine if the intervention had any impact on the development of self-assessment skills of the interpreters.

### 3.2 Respondents

A sample representative of the population deemed as ‘professional interpreters’ according to the above stated definition was selected as the unit of analysis for the research. For the purpose of the study, the term ‘professional interpreter’ was defined as an interpreter presumed to not simply be competent but having mastered their skill with prior experience and/or training in interpreting and adhering to high standards of conduct supported by a code of practice. Interpreters who, at the time of the study, were employed full-time within the Interpreting Unit of the Language Services Section at the National Parliament of the Republic of South Africa were chosen as the sample population for this study.<sup>6</sup> A week after the link to the questionnaire was sent out via email, ten of the interpreters

---

<sup>5</sup>The necessary ethical clearance was provided from Stellenbosch University as well as institutional clearance from the Parliament of SA. The National Health Research Ethics Committee (NHREC) registration number is REC-050411-032.

<sup>6</sup>Regarding the recruitment policy of parliament for interpreters, it is suffice to state here that as there is relative short history of interpreting, interpreters were initially recruited from the teaching profession. However, over the last couple of years some inroads have been made as

in the Unit completed the questionnaire and were subsequently divided into the control group (5 respondents) and the experimental group (5 respondents) using the quota matching method. The characteristics which were used to divide the respondents equally into the two different groups were;

1. working languages;
2. interpreting education and/or training and
3. experience in interpreting.

### 3.3 The questionnaire

The questionnaire was divided into different sections, all of which aimed to collect data on:

- The biographical details of the respondents' pertaining to their experience and education in interpreting
- The perceived knowledge of the respondents as pertaining to their self-assessment activities and their awareness of his/her strengths and weaknesses in interpreting performance and lastly
- The respondents' perceived knowledge about the evaluation process and the applicable criteria considered when evaluating an interpreting performance

A copy of the questionnaire in its entirety is included in Addendum B. The main section is discussed:

Question 1 of the questionnaire dealt with the working languages of the interpreters and the researchers asked the question to determine the working languages. The majority (80%) of the respondents indicated that they provide interpreting in English.<sup>7</sup> The data for the Language A<sup>8</sup> distribution was indicated as (20%) Afrikaans, (20%) isiZulu, (20%) SiSwati, (10%) isiNdebele, (10%) Sepedi, (10%) Sesotho and (10%) Tshivenda.

---

trained interpreters were appointed. These interpreters currently provide for all the 11 official languages as well as Sign Language. Please see Lesch2010 for information on recruitment and training of parliamentary interpreters.

<sup>7</sup>Although only 80% of the respondents indicated they deliver interpreting services in English, it forms part of the employment contract of the interpreters in Parliament that they must all be able to interpret into English as their B language.

<sup>8</sup>Language A is representative of the respondents' mother tongue or first language.

LangSci:  
Create  
proper  
link

Questions 2, 4 and 5 of the questionnaires dealt with the interpreters' practical experience in interpreting. The questions sought to determine how many years' experience the interpreter had, if the interpreter had any experience in interpreting before they started interpreting at Parliament and lastly in what setting (e.g. court, health care, conference), the interpreter had experience. The majority (40%) of the respondents indicated they had between 5–9 years' experience as an interpreter, followed by (30%) indicating they have between 10–20 years' experience as an interpreter. Two respondents (20%) indicated that they had less than 5 years' experience while only one respondent (10%) indicated that they had more than 21 years' experience in interpreting.

In question 4, the majority (70%) of the respondents had indicated that they had prior experience in interpreting before they started interpreting at Parliament.

Question 5 was an open-ended question inquiring as to the setting where the respondent had provided interpreting services. All 7 respondents who had indicated prior experience responded to the question and the text responses were categorized as follow; three interpreters indicated that they had conference interpreting experience. This included interpreting for the Truth and Reconciliation Commission, Provincial Legislature, conferences, general meetings and workshops. Two interpreters indicated that they had court interpreting experience. One interpreter indicated they had experience in educational interpreting at university. One interpreter indicated that they had their own company which provided interpreting services, while another interpreter had indicated that they had been working as a freelance interpreter for 4 years. In both these instances the specific setting where interpreting services were rendered were not provided.

Question 3 dealt with the interpreters' employment at Parliament. The researchers wanted to determine how many years' the interpreter had been interpreting in the environment of the Parliament. This data would also indicate how experienced the interpreter is in conference interpreting particularly with Parliamentary speeches and terminology. The responses indicated that 50% of the respondents had been working as an interpreter in Parliament for 5–9 years. 20% had been working for 10–20 years and 30% had been working for less than 5 years.

The majority (80%) of the respondents indicated that they held a qualification in interpreting, translation or a language practitioner related qualification. Of these, four indicated that they held a tertiary diploma; one indicated a Bachelor's degree and three indicated that they held an Honours degree, i.e. a qualification after the BA degree that gives access to study on Masters level. The majority (60%) of the respondents indicated that they received informal training. The informal



training was listed as in-house training and short courses.

### 3.4 Data collection

The data collected from the biographical information in the questionnaires was used in the quota matrix matching method from experimental studies to divide the respondents equally into the control group (five respondents) and the experimental group (five respondents). A quota sampling method entail the gathering representative data from a group. As opposed to random sampling, quota sampling requires that the individuals are chosen out of a specific subgroup.

The respondents were all coded by using letters of the alphabet (A–J) and the data obtained from the biographical information pertaining to prior experience and education in interpreting were then tabulated according to these codes and the matching method was used to divide the respondents on a random basis equally into two groups; namely the experimental group that will be exposed to training, and the control group that won't receive any training.

Table 1: Matching method division of respondents

	2 Experience	4 Prior Experience	5 Setting	6 Qualification	7 Type of Qualification	8 Informal training	9 Type of training
A	21+	Yes	TRC	No		No	
B	1–5	No		Yes	Hons.	Yes	In-house
C	10–20	No		Yes	Postgraduate Diploma	Yes	In-house
D	5–9	Yes	Church	No		No	
E	5–9	Yes	Freelance	Yes	Postgraduate Diploma	Yes	In-house
F	10–20	Yes	Freelance	Yes	Bachelors	Yes	Practical
G	5–9	Yes	Church	Yes	Hons.	No	
H	1–5	No		Yes	Diploma	Yes	Short Course
I	10–20	Yes	Court/ Conferences	Yes	Postgraduate Diploma	Yes	In-house
J	5–9	Yes	University/ Legislature	Yes	Hons.	Yes	In-house

The second part of the data collection was the experiment itself. The experiment comprised three major pairs of components:

1. Independent (*Black Box*) and dependent variables (self-assessment skills)
2. Pre-testing and post-testing
3. Experimental (participate in four self-training sessions on *Black Box*) and control groups.

Table 2: Experimental group and Control Group

Experimental Group	Control Group
B	A
C	E
D	F
G	H
I	J

The experimental stimulus, *Black Box*, was administered to the experimental group over a course of four sessions. The respondents from the experimental group were required to complete four self-assessment sessions of between 30–60 minutes each (consisting of 10–13 minutes’ introduction and interpreting exercise itself; 12–15 minutes listening to the recording of target and source speech and finally 13–16 minutes’ completion of the self-assessment grid). The self-training sessions consisted of one simultaneous interpreting exercise on the software *Black Box*, where a parliamentary speech of between 6 and 8 minutes had to be interpreted. The speech was made up of a video as well as audio clip. The target (interpretation of respondent) and source (original text) speeches of the self-assessment sessions were recorded on *Black Box* which compresses the target and source speech into one single audio file. After completing the interpreting exercise, respondents were required to listen to the recording and use a provided grid for self-assessment. The self-assessment grids were collected after each respondent had completed it. The aggregate out of 15 for each session was recorded to track progress and compare the marks from each session.

The ten respondents who participated in the experiment interpreted into their A language, i.e. 7 different languages which formed part of the experimental output. As indicated in section 3.2.1, the data for the Language A distribution was indicated as (20%) Afrikaans, (20%) isiZulu, (20%) SiSwati, (10%) isiNdebele, (10%) Sepedi, (10%) Sesotho and (10%) Tshivenda. An expert for each of these languages were utilized to conduct the expert rating. These experts are rated on their seniority in terms of their language specific background but also their interpreting experience as well as the in-house principles that applied in Parliament. They made use of the same assessment grid as utilised by the participants and as attached in the Addendum A.

The video material used in the self-assessment interpreting sessions were recorded during National Parliamentary sittings of the National Assembly, readily

Which  
section?

#### 4 Experimenting with computer-assisted interpreter training tools

available on the National Parliament of the Republic of South Africa's YouTube channel. The video material consisted of four speeches from different debates and different political parties and the length varied from 6 to 8 minutes. The dominant language spoken in the recordings was English, but session 2 contained some Setswana and isiNdebele and session 4 contained some isiZulu. Table 3 below is a representation of the four different sessions.

Table 3: Summary of the different interpreting sessions

	Session 1	Session 2	Session 3	Session 4
Language(s)	English	English Setswana isiNdebele	English	English isiZulu
Topic	Debate on the State of the Nation 2016	Question Session	Debate on the State of the Nation 2016	Debate on the Marikana Commission of Inquiry
Duration	6:35min	7:45min	7:47min	6:35min

For the first self-assessment session on *Black Box* the respondents allocated a mark for pre-testing, i.e. before the intervention. In the final self-assessment session the respondents in the experiment allocated a mark that represent their post-testing mark.

In a traditional experiment the control group is never exposed to the stimulus. However, for this research the only way of obtaining a recording which was similar to that of the experimental group, the control group had to be exposed to the stimulus. The pre- and post-testing for the control group was done using the same instrument, *Black Box*, which was the experimental stimulus in this study. Thus, unlike a traditional experiment in which the control group is never exposed to the stimulus, the control group in this experiment received exposure to the stimulus as they had to complete one session of self-assessment of interpreting performance on *Black Box* in order for their interpreting assessment score. This was done by allowing the control group to listen to both the target and source texts and to conduct self-assessment. Self-assessment does not necessarily require one to listen to the recording, however in this case it was expected from interpreters to listen to it and reflect on it for self-assessment. The mark obtained in the one session completed on *Black Box* by the control group was used as data for pre- and post-testing. Thus, the comparison between the two groups includes the fact that the experimental group received exposure and training on *Black Box* over four sessions, whereas the control group only had the one session

on *Black Box*.

At the end of the four sessions, the respondents from the experimental group were interviewed, which aimed to extend the findings of the quantitative data. The main aim of having the interviews was to conduct a follow-up and evaluate the perceptions of the respondents regarding their strengths and weaknesses as well as the criteria they used when evaluating interpreting performance post experiment. The interviews were structured and based on written questions (see Addendum 5) and were conducted individually after the respondents from the experimental group completed their final self-assessment session.

This link  
needs  
fixing

## 4 Results

Hypothesis testing was used to analyze the quantitative data obtained from the experiment. In hypothesis testing two opposing hypotheses are measured. The two hypotheses are known as the null hypothesis and the alternative hypothesis. The alternative hypothesis is based on the aim of the research, in other words, that the observed differences are the result of real effects, while the null hypothesis would state that there is no significant difference between the populations specified by the study. In hypothesis testing, the null hypothesis is assumed to be true. In this instance, the null hypothesis would be that there is no difference between the means from the absolute errors pre- and post-experiment from the experimental group. The alpha of 0.05 is used as a guideline to determine to what extent the hypothesis may be accepted or rejected. In most analyses, an alpha of 0.05 is used as the cut-off for significance. If the p-value is less than 0.05 ( $p < 0.05$ ), the null hypothesis is rejected. If the p-value is larger than 0.05 ( $p > 0.05$ ) the null hypothesis is accepted to be true.

why  
state the  
mathe-  
matical  
format?

Against this background the hypothesis regarding the effectiveness of training software as a self-assessment skills development tool for practicing interpreters are evaluated.

### 4.1 Was there a difference in the correlation of self-assessment ratings from the experimental group and the ratings from the expert assessor post-experiment?

As hypothesis testing was used to analyze the quantitative data obtained from the experiment, the hypothesis test was set up to determine the validity of the statistical claim that there was no difference between the absolute error means pre- and post-experiment. The p-value from the experimental data was calculated

at 0.24198 (see Figure 1) which meant that based on the  $p$ -value, a significant difference could not be concluded. Alternatively, it can be stated as in Table 4.

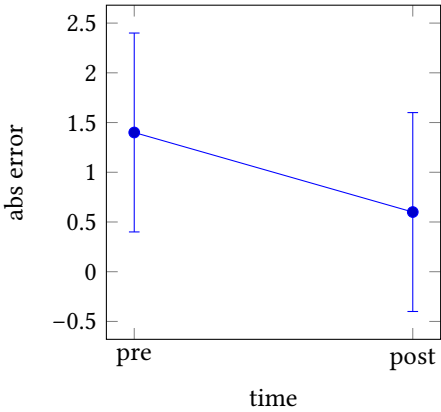


Figure 1: Means of absolute errors pre-experiment and post-experiment

Please confirm graph

Table 4: Descriptive statistics for hypothesis testing

Descriptive statistics			
Effect	Level of factor	Number	Absolute error mean
Total		10	1.00
Time	Pre	5	1.40
Time	Post	5	0.60

Although the results were not statistically significant, the descriptive statistics did indicate that over time the experimental group’s absolute error mean ratings did decrease (see Table 5). The decreasing absolute error indicates that after exposure to the experiment there were more self-ratings which corresponded with the rating from the experts. The comparison of pre-experimental and post-experimental data (see Table 6) pertaining to the self-ratings and expert-ratings from the experimental group indicated that, pre-experiment, only one respondent could accurately rate themselves in accordance with the rating by the experts. However, there may be shortcomings as this is a small-scale experiment and data was only collected at two instances. Post-experiment data indicated that

Table 5: Experimental group ratings pre- and post-experiment

Experimental Group					
Pre-experiment			Post-experiment		
Self-rating (out of 15)	Expert Rating (out of 15)	Difference	Self-rating (out of 15)	Expert Rating (out of 15)	Difference
11	11	0	12	12	0
9	12	3	11	11	0
12	11	-1	12	12	0
13	12	-1	13	12	-1
13	11	-2	13	11	-2

three respondents could accurately rate themselves in accordance with the rating by the experts.

#### 4.2 Was there a difference in the self-assessment ratings of the control group when compared to the experimental group post-experiment?

The means between the final sessions from the experimental group and the control group (see Table 6) did indicate a difference, with the experimental group scoring higher ratings overall. The experimental group's average final self-assessment ratings were calculated as a mark of 12.2 out of 15 and the average final self-assessment rating from the control group was calculated as a mark of 10.8 out of 15. However, since the control group only had one set of ratings – it could not be used for statistical analysis. The possibility exists that there are other variables which may have contributed to the difference in ratings.

#### 4.3 Do the self-assessment sessions give the interpreters a better awareness of their strengths and weaknesses in interpreting?

The analysis of the qualitative data from the questionnaire indicated that it was the perception of the majority of respondents that their strengths in interpreting far outweigh their weaknesses. The qualitative data from Question 11 (*How often do you struggle with the following challenges in interpreting?*) indicated that it was the perception of the majority of the respondents (80%) that they seldom struggled with challenges in interpreting.

Table 6: Comparison of ratings for experimental and control group

Experimental Group					
Pre-experiment			Post-experiment		
Self-rating	Expert Rating	Difference	Self-rating	Expert Rating	Difference
12	12	0	12	12	0
11	11	0	12	9	-3
12	12	0	7	8	1
13	12	-1	13	12	-1
13	11	-2	10	9	-1
Averages of groups					
12.2	11.6		10.8	10	

The qualitative data from Question 12 (*Indicate your ability with regard the following in simultaneous interpreting*) indicated that there was a positive perception among the majority of respondents when asked a negative Likert-scale question; for example, when the question was posed in the negative, the majority of answers were found among the choices of “*never*” and “*seldom*”. When posed with a positive question, the majority of the answers were “*frequently*” and “*always*”. It was seldom that a respondent indicated a challenge or weakness in their interpreting performance.

The marks obtained, both the self-assessment rating as well as the ratings from the experts, in the self-assessment grids from the experimental group respondents were high (see Table 6). However, the specific questions posed under each macro error section of the self-assessment grids showed that the respondents did encounter challenges in their interpreting performance, especially when it pertained to the interpretation of idiomatic expressions and accurate interpretation of numbers and dates.

In the qualitative data from the interviews it was the perception of all respondents that the self-training sessions gave them a better awareness of their strengths and weaknesses. In answer to Q3 (*Were you satisfied with your interpreting performance?*), four of the respondents were satisfied with their interpreting performance and one respondent indicated that they were “not quite” satisfied. In answer to Q4 (*Was your interpreting performance better or worse than you expected?*), all the respondents indicated that their performance was better than they had expected. In answer to Q5 (*In the questionnaire there was a section per-*

*taining to your abilities in interpreting. After having conducted self-assessment – do you think that your initial judgments were correct?*), two of the respondents indicated that their initial judgments of their abilities in interpreting had been correct. One respondent indicated that their ability was better than they had expected. Other respondents indicated that their judgments were correct but that they “*can do better of course*”. One respondent indicated that “*some things were better than I thought they would be*”. In answer to Q14 (*Do you feel that the self-study sessions has given you a better awareness of your strengths and weaknesses in interpreting?*) all respondents indicated that the self-assessment session gave them a better awareness of their strengths and weaknesses.

#### **4.4 Do the self-assessment sessions give the interpreters a better awareness of the criteria used in the evaluation of interpreting performance?**

The data collected from the questionnaires indicated that the perception of the respondents regarding the criteria used in the evaluation of interpreting performance was quite vague and incomplete. Question 13 of the questionnaire is an open-ended question inquiring from the respondents to list the criteria they find important in the evaluation of an interpreting performance. The data collected from this question was arranged in tabular format according to the macro errors; 1) accuracy, 2) target language and 3) delivery (see Figure 2; also see Addendum 5 again). Under each of the macro errors examples of errors were listed. A heading for ‘other’ was been added. From the data provided, it was deduced that the respondents were not completely aware of the criteria used when evaluating an interpreting performance. Only half of the respondents (50%) indicated that accuracy is important in the evaluation of interpreting performance, while only 20% of respondents indicated that target language is important in the evaluation of interpreting performance. The majority of examples listed by the respondents were found under the macro error of delivery. However, each respondent listed only one item under this macro error.

The data collected and analyzed for the macro error of accuracy indicate that it is the perception of half of the respondents (50%) that accuracy is important in the evaluation of interpreting performance. The terms “accuracy”, “content accuracy” and “message accuracy” were used. None of the respondents list “omissions” or “additions” as criteria. There were also no examples given of what constitutes “accuracy”.

The data analyzed for target language indicate that it is the perception of 20%



#### 4 Experimenting with computer-assisted interpreter training tools

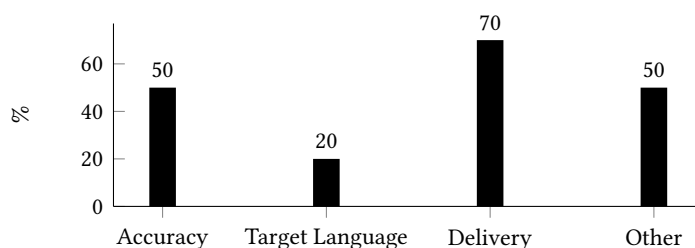


Figure 2: Criteria used for evaluation of interpreting performance: Distribution of perceptions of respondents

of respondents that target language is important in the evaluation of interpreting performance. Only two respondents (20%) listed criteria pertaining to the category of target language by indicating that ‘terminology accuracy’ and ‘vocabulary’ are important criteria in the evaluation of interpreting performance. One respondent (10%) indicated that ‘sentence construction’ is important when evaluating interpreting performance.

In the analysis of the data under the macro error of delivery, seven respondents (70%) listed criteria which pertains to delivery. Eleven different micro errors were listed as criteria important in the evaluation of interpreting performance. The analysis of data gathered from this question reveals a strong focus on the macro error of delivery when seen in relation to the variety of micro errors listed. Two respondents (20%) listed the micro error pertaining to tone of voice by stating: ‘tone of voice follows the speaker’ and ‘voice tone’. Two respondents (20%) listed criteria pertaining to the micro error of audibility by listing: ‘audibility’.

Eleven other micro errors pertaining to the category of delivery were listed; Absence of fillers – 10%; Avoiding long pauses – 10%; Breathing – 10%; Consistency – 10%; Coherence – 10%; Correct intonation – 10%; Delivery smooth and clear – 10%; Pleasant to hear presentation – 10%; Time lag 10%. Only one respondent (10%) listed criteria across all three different macro errors (accuracy, target language, delivery).

In the qualitative data from the interviews, it was the perception of all respondents that the exposure to the self-assessment sessions had improved their understanding of the criteria used in the evaluation of an interpreting performance. In answer to Q6 (*Before the self-study sessions – were you aware of criteria used in the evaluation of interpreting?*), two of the respondents indicated that, before the self-assessment sessions, they were not aware of the criteria used in the evaluation of interpreting. Three respondents indicated that they were aware of the criteria

used in the evaluation of interpreting. In answer to Q7 (*Do you feel that your understanding of the criteria has improved with the self-study sessions?*) when asked if the respondents' understanding of the criteria had improved with the exposure to the self-assessment sessions, there was a consensus among all five respondents that their understanding had improved. One respondent indicated that their understanding of the criteria improved, especially after completing the electronic questionnaire. The questions posed in the questionnaire might have triggered the respondents' thoughts and lead them to reflect on the criteria used when evaluating interpreting performance. The self-assessment grids which were used in the self-assessment session also clearly indicated the various macro errors and criteria used for the evaluation of an interpreting session.

## 5 Conclusion

One need to emphasize that the study did not seek to evaluate the performance of the interpreters but it was rather aimed at evaluating self-assessment skills of the interpreters. The marks from both the self-assessment and expert ratings were relatively high – indicating that the interpreters perform at a high level. This ought to be expected since the interpreters are no longer student interpreters, but full-time professional interpreters. The means (both the self-ratings and the ratings by the experts) between the final sessions from the experimental group and the control group did indicate a difference in the averages from the groups (see Table 6), with the experimental group scoring higher ratings. This indicates that their self-assessment did differ from that of the control group. However, there are several variables which may have contributed to this difference in ratings.

The empirical study sought to obtain quantitative and qualitative data. The primary research aim of the study set out to evaluate whether the CAIT tool, *Black Box*, was effective in the development of self-assessment skills in professional interpreters. The primary research aim was sub-divided into four research questions and that was covered addressed above under the results. According to the results of this small scale research study, it can be concluded that the CAIT tool, *Black Box*, may prove effective in the development of self-assessment skills in professional interpreters.

Langsci: Rename Addendum => Appendix

## Addendum A: An example of the self-assessment grid

SELF-ASSESSMENT SHEET					
<b>Debate on the State of the Nation Address</b>					
Duration: 6:35min					
Read through all the questions in this self-assessment sheet before you start with the playback of the recording. You are allowed to pause, rewind and make notes while listening to the recording.					
PARTICIPANT CODE:					
<b>1. ACCURACY / CONTENT OF MESSAGE:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Omissions, Additions, Accuracy <i>The interpreter must convey the message in a complete, correct and intelligible manner in the target language.</i>					
1.1 Was important information omitted in this interpreting session?					
				YES	NO
<b>2. TARGET LANGUAGE</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Vocabulary, Sentence Construction, Idiomatic language use, Grammar <i>The interpreter must always use the most appropriate vocabulary and be loyal to the register of the speaker.</i>					
2.1 The following idiomatic language was used in the speech					
<ul style="list-style-type: none"> <li>• write down how each statement was interpreted</li> <li>• comment on whether the phrase was interpreted into idiomatic target language</li> </ul>					
[00:20mins] "anxious coin tossing"					
[1:40mins] "He spoke a lot today about iron and steel. Well, let me tell you something: When it comes to the ANC, <b>they iron over the problems and steal all the money.</b> "					
<b>3. DELIVERY / COHERENCE / TECHNIQUES and PRESENTATION</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Inarticulate speech, Pauses and hesitations, Audibility, Fillers <i>The interpreter must maintain sufficient speed to convey the full message of the speaker, employing mechanisms to cope with various complexities, remaining clear and concise.</i>					
3.1 Is the interpreting audible / clear?					
				YES	NO
3.2 Are there any fillers (uhm, ah)?					
				YES	NO
3.3 Are there any unfinished sentences?					
				YES	NO
3.4 Are there any strange noises (coughing, sighing, heavy breathing)?					
				YES	NO
3.5 Is the intonation natural or monotonous?					
				NAT.	MON.
3.6 Is the lag-time managed well?					
				YES	NO
<b>TOTAL MARK:</b>					<b>/ 15</b>

## Addendum B: Questionnaire

### INTERPRETING QUESTIONNAIRE

Thank you for taking part in this study! This questionnaire is completed anonymously.

#### Section A: Interpreting Experience

**1** In what language(s) do you provide interpreting?

**2** How many years' experience do you have as an interpreter?

21 + years

10–20 years

5–9 Years

Less than 5 years

**3** How long have you been employed as an interpreter in Parliament?

21 + years

10–20 years

5–9 Years

Less than 5 years

**4** Did you have experience in interpreting before you started working at Parliament?

Yes

No

**5** If yes, please specify where you have rendered interpreting services (example; court, clinic, any other):

**6** Do you hold a qualification in interpreting/ translation or language practice?

Yes

No

**7** If yes, what qualification do you hold?

Diploma

Bachelors

Masters

Honours

PhD

Other, please specify:

**8** Have you received any informal interpreter training?

Yes

No

**9** If yes, please specify what type of training you received (example; short courses, in-house training, any other):

### Section B: Self-assessment activities

10	How often do you...	Never	Seldom	Frequently	Always	N/A
10.1	Record your interpreting sessions					
10.2	Listen to recordings of your interpreting sessions					
10.3	Take note of terminology which is challenging in an interpreting session					
10.4	Take note of challenges presented in an interpreting session					
10.5	Conduct self-assessment on an interpreting performance					

### Section C: Interpreting strengths and weaknesses

11	How often do you struggle with the following challenges in interpreting?	Never	Seldom	Frequently	Always	N/A
11.1	Interpreting proper names					
11.2	Interpreting numbers and figures					
11.3	Interpreting dates					
11.4	Understanding the speakers' accent					
11.5	Following the speakers' speed					
12	Indicate your ability with regards to the following in simultaneous interpreting:	Never	Seldom	Frequently	Always	N/A
12.1	I struggle to provide an accurate message					
12.2	I pause within the middle of a sentence					
12.3	I struggle with target language register					

		Never	Seldom	Frequently	Always	N/A
12.4	I struggle with target language terminology					
12.5	I hesitate					
12.6	I have a monotonous intonation					
12.7	I use filler words such as uhm and ah within a sentence					
12.8	My speech is unclear					
12.9	I struggle with target language grammar					
12.10	My target language use is unidiomatic					
12.11	I omit information					
12.12	I add information					
12.13	I do not finish sentences					
12.14	My message delivery is incoherent					
12.15	I struggle with microphone use					
12.16	I need to improve my simultaneous interpreting technique					
12.17	I struggle to concentrate while interpreting					
12.18	I speak too fast					
12.19	I breathe loud					
12.20	I get emotionally involved					
12.21	My delivery is smooth and flows with ease					
12.22	I convey the message accurately					
12.23	I do not make irritating noises					
12.24	My voice sounds pleasant					
12.25	I use the appropriate terminology					
12.26	I do not stop in the middle of a sentence					

#### **Section D: Evaluation of Interpreting Performance**

**13** List the criteria which you find important in the evaluation of an interpreting performance:

**14** Define each of the following macro errors when used to evaluate interpreting performance:

14.1 Accuracy

14.2 Delivery

14.3 Target Language Quality

**15** Provide examples of errors according to the following macro errors (for example; accuracy = omissions):

15.1 Accuracy

15.2 Delivery

15.3 Target Language Quality

**15** Do you have any other comments?

## **Addendum C: Interview questions**

- Q1:** Was this the first time you recorded your interpreting performance?
- Q2:** Was this the first time you listened to yourself interpreting?
- Q3:** Were you satisfied with your interpreting performance?
- Q4:** Was your interpreting performance better or worse than you expected?
- Q5:** In the questionnaire there was a section pertaining to your abilities in interpreting. After having conducted self-assessment – do you think that your initial judgements were correct?
- Q6:** Before the self-study sessions – were you aware of criteria used in the evaluation of interpreting?
- Q7:** Do you feel that your understanding of the criteria has improved with the self-study sessions?
- Q8:** In your first self-study session and self-assessment did you find it difficult to assess yourself?
- Q9:** Do you feel that the self-study sessions have developed your self-assessment skills?
- Q10:** Do you feel that it is easier being assessed on your interpreting performance by someone else?
- Q11:** Do you think that if someone was to assess this very same assessment that you would receive the very same mark?
- Q12:** Did you find the self-study sessions useful in order to conduct self-assessment?
- Q13:** Did you find the self-assessment grids useful in your self-assessment?
- Q14:** Do you feel that the self-study sessions have given you a better awareness of your strengths and weaknesses in interpreting?
- Q15:** Do you feel that the self-study sessions have improved your interpreting performance?
- Q16:** What did you find most useful in the self-study sessions?
- Q17:** Do you feel that the self-study sessions have made you more confident in conducting self-assessment on your interpreting performance?
- Q18:** Do you have any other comments?



## Chapter 5

# Technologies and role-space: How videoconference interpreting affects the court interpreter's perception of her role

Jerome Devaux

The Open University

Back in 2000, videoconference (vc) systems were introduced in criminal courts in England and Wales so that defendants could attend their pre-trial court hearings from prison. Since then, the number of cases heard via vci technologies has been on the increase. In order to be able to conduct a hearing remotely, courts and prisons are equipped with cameras, screens, microphones, and loud-speakers which link up both locations so that participants can hear and see each other. In terms of research, various reports on the viability of such systems acknowledge the benefits of conducting court hearings remotely, whilst also highlighting shortfalls. Interestingly, most of these studies were carried out in a monolingual setting, and fewer studies examine the impact of vci equipment in multilingual court settings. In this context the interpreter's role, and more particularly her role perception when technologies are used in a courtroom, remains under-explored.

This paper focuses on one specific type of interpreting technologies, namely Videoconference Interpreting (vci). It aims to report on a doctoral study that explores practising court interpreters' perceptions of their role in England and Wales, when they interpret through vci systems. To do so, semi-structured interviews were conducted with eighteen participants, and the data gathered was analysed through the prism of role-space. It will be argued that, in line with Llewellyn-Jones2014's (Llewellyn-Jones2014) models, some interpreters perceive their role as a 3D fixed entity, whilst other create a 3D continuum. However, building on Llewellyn-Jones2014, this paper will also demonstrate that, unlike in face-to-face court hearings, technologies force some interpreters to create split role models.

Please shorten this abstract such that the main text begins on the title page



## 1 Introduction

Since the late 1990s, videoconference (vc) systems have been used in criminal courts in the UK (Plotnikoff1999; Plotnikoff2000) as a means to reduce cost, enhance security, and speed up proceedings. According to (Braun2016b), 90% of Magistrate's Courts and all Crown Courts were equipped in 2013 with the necessary vci equipment to enable courts to establish an audio and video feed between the participants physically present in a courtroom and the defendant or witness attending from a remote location. In other words, defendants can attend their own court hearing without leaving the prison where they are incarcerated.

Such systems are also used in multilingual court hearings in the UK. In this context, the interpreter can be co-located with the participants in court or with the remote defendant or witness. Braun2011a makes a useful distinction by referring to Videoconference Interpreting (vci) A, where the interpreter is in the courtroom and vci B, where the interpreter is co-located with the remote defendant or witness.

Research in the use of vci systems in courts dates back to the 1990s and was characterised by its primary focus on monolingual court settings. More recently, the valuable studies carried out as part of the Avidicus<sup>1</sup> projects have filled in parts of the research void on the use of vci equipment in multilingual legal settings. Although these projects and other research cover many different grounds, the interpreter's perception of her role<sup>2</sup> remains largely unexplored.

This article builds on the current body of knowledge in Interpreting Studies (is) by examining eighteen interviews carried out with practising spoken language court interpreters in England and Wales. Their interviews are analysed through the medium of role-space, a relatively new theoretical framework developed by Llewellyn-Jones2014.

The first two Sections briefly review the literature on the use of vci equipment in court and the interpreter's role. §3 summarises the methodology used, and §4 analyses the data gathered. Finally, §5 discusses the data in light of the findings from the literature review and formulates recommendations for interpreter training. It is posited that the use of technology enhances and/or creates different factors that can affect various aspects of the interpreter's role-space.

---

<sup>1</sup>Avidicus stands for Assessment of Videoconference Interpreting in the Criminal Justice Service.

<sup>2</sup>For purely stylistic reasons, the term 'interpreter' will sometimes be replaced by the feminine personal or possessive pronouns 'she' or 'her', whereas he/his/him will refer to either the defendant or the witness.

## 2 Videoconference interpreting in court settings

The use of vci equipment in a legal setting was examined as early as the 1990s. Research at the time was mainly restricted to monolingual court settings, with a specific focus on the us court context (**Radburn-Remfry1994; Thaxton1993**). From then onwards, research in a monolingual setting has evolved mainly around three intrinsically related areas. First, scholars such as **Johnson2006** question the legality of using vci technology to mediate a court hearing as it could infringe the defendant's right to due process. Similarly, **Radburn-Remfry1994** and **Thaxton1993** raise concerns as regards the impact that vci equipment could have on the fairness of the court proceedings. The second main research theme focuses on the impact that vci equipment has on participants' perceptions of the court hearing, and how court participants interact. Studies reveal that it is more difficult to assess emotions (**Radburn-Remfry1994**), body language (**Fullwood2008**), and a witness's credibility (**Roth2000**). There is also a risk that participants feel detached from the process (**McKay2016**), and the working relationship between the remote defendant and the participants in court may be questioned (**Hodges2008**). **Verdier2011** also demonstrate that the conversation can be more fragmented, utterances can overlap (**Licoppe2015**), and the defendant may be more reluctant to interact (**Licoppe2014**). Finally, a number of studies have discussed technological issues, including the impact on interaction of poor sound and video quality (**Haas2006; Plotnikoff2000**). Another research area also seems to emerge as studies such as **Licoppe2013** no longer investigate quality-related issues in terms of equipment, but they examine how the interaction itself is produced (e.g.: who orchestrates the camera moves, and how this affects the interaction).

In multilingual settings, studies on the use of vci equipment form part of a relatively new research area, which has been primarily examined within the realm of the Avidicus projects. This research investigates the use of vci and Remote Interpreting (RI) in various legal settings such as criminal courtrooms and police stations across Europe. They also offer training guidelines and recommendations to various legal stakeholders. Their studies are quite far reaching, and some of their findings confirm those in a monolingual setting. For instance, Avidicus 1 reveals that it is more difficult to establish a rapport with the participants on the other side of the screen (**Rombouts2011**). It also demonstrates that vci requires more synchronisation in terms of interaction and turn-taking, and it is more conducive to overlapping turns and artificial pauses (**Balogh2011**). Furthermore, interpreters reported that they found it more stressful, isolating, and tiring (**Miler-Cassino2011**). In a bid to further explore the impact of vci in a legal

setting, Avidicus 2 establishes a list of interrelated factors that affect interpreting quality and a list of strategies developed by interpreters. It also offers some strategies to interpreters to overcome issues relating to the use of vci. Finally, Avidicus 3 takes stock of the use of vci equipment in twelve European countries and, for each of them, the findings are thematised under nine areas: procurement, equipment and maintenance, uses, participant distribution, pre-vc/post-vc, mode of interpreting, vci management communication management, and working arrangements. According to **Braun2016a**, it transpires that in England and Wales there are various vci equipment suppliers operating, and vci equipment is fitted within existing courtrooms, which can dictate the position of the screen and cameras. These set-ups can lead to various potential layouts and create different constraints. Furthermore, vci hearings tend to be rather short, and are characterised by a lack of pre-briefing and debriefing sessions. Finally, the interpreter works mainly in consecutive mode, and the issues of rapport building with remote participants are further highlighted.

Other studies have been conducted in the area of vci-mediated legal interpreting outside the realm of the Avidicus projects. For instance, **Ellis2004** examines the fairness of its use in refugee hearings in Canada. This report confirms that the use of vci leads to a more impersonal means of communication, and it also highlights technical issues regarding poor audio and video quality. Similarly, the Bail for Immigration Detainees and the British Refugee **Bail2008's** (**Bail2008**) report draws similar conclusions, and it also reveals that the use of vci equipment distorts body language in immigration hearings. Furthermore, in English criminal courts, **Fowler2012** examines the use of equipment, the interpreter's working conditions, and the interaction management. Her studies show that the interpreter is a more visible court actor when vci equipment is used. Finally, in a recent study **Devaux2017a**, I investigated the interpreter's ethical rationalisation process and argued that interpreters rationalise ethical dilemmas mainly through their codes of ethics. However, specific ethical issues arise in vci A and/or vci B, for which other ethical paradigms, such as consequentialism or virtue ethics, need to be considered.

Overall, research carried out in vci tends to focus on various paradigms that evolve around the use of vci equipment and its impact on the interaction. Results show that there are similar difficulties in a mono- or multilingual context, be it related to technical difficulties or interaction management. Interestingly, the legality concerning the use of vci is not as prominent a research area as it is in monolingual settings. Based on the literature review, it is also striking that the court interpreter's role in vci is an underexplored research area, especially as

this theme has been studied widely in various face-to-face contexts.

### 3 The court interpreter's role

The interpreter's role has been examined in many different public service settings, which has led to many role labels being coined. To name but a few, interpreters have been referred to as a conduit, a clarifier, a culture broker or an advocate (Niska2002); a filter, a detective, a multi-purpose bridge, a diamond connoisseur, or a miner (Angelelli2004); and a helper, a social worker, an advisor or an advocate (Grbic2011).

Building on the seminal work by Wadensjö1998, the body of research demonstrates that the interpreter can adopt different role labels during the same Interpreter-Mediated Event (IME). During psychotherapeutic sessions, Bot2009 for instance, describes the role of the interpreter as a continuum where the interpreter as a conduit and the interpreter as an active participant are situated at either end of such a continuum. Similarly, Mason2009 argues that the interpreter is an active member in immigration interviews, and her positionings will change and adapt in light of other participants' responses.

When examining the role of the court interpreter more specifically the literature reveals that, contrary to the ideology often imposed by the court, the interpreter is a conduit (Laster1994), and she can also adopt several roles during a court hearing (Berk-Seligson1990; Martin2009). Similarly to other public service settings, many labels have been created to identify her role. For instance, Hale2008 observes that the court interpreter can be: an advocate for the minority language speaker, an advocate for the institution of the service provider, a gatekeeper, a facilitator of communication, and a faithful renderer of the other's utterances. Other researchers describe the court interpreter's role as an impartial translation machine, a linguistic and cultural bridge, an expert witness (Mikkelsen1998), a cultural or linguistic mediator, or a communication facilitator (Nartowska2016).

The above studies rely on attributing a role label with certain characteristics to the role(s) that researchers observe or analyse. However, Gentile1996 state that a "kaleidoscope of role (...) is not conducive to the creation of a professional identity", and one could question the extent to which creating different role labels with sometimes blurred characteristics can contribute to a professional identity. A potential means to circumvent the creation of more role labels may reside in the use of role-space, a rather new theoretical framework in is that became more widely-known due to Llewellyn-Jones2014's (Llewellyn-Jones2014) pub-

lication. Role-space is based on the three-dimensional conceptualisation of the interpreter's role alongside three axes. First, the z-axis, Presentation of Self, refers to the interpreter herself, and how much or how little information she provides about herself during an IME. The x-axis, Participant Alignment, indicates whether she is siding more towards one party, or whether she remains neutral. Finally, the y-axis, Interaction Management, indicates the extent to which she manages the interaction between the parties. Figure 1 summarises the template that is used in this study. Worth noting is the fact that a role-space model is organic and that the interpreter's presentation of self, participant alignment, and interaction management may fluctuate alongside their respective axes in order to reflect changes within an IME.

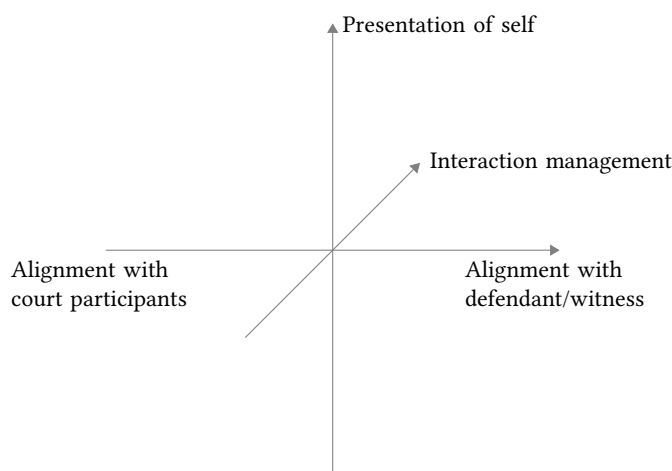


Figure 1: Llewellyn-Jones2014's (Llewellyn-Jones2014) role-space template

In order to design the interpreter's role-space, Llewellyn-Jones2011 (Llewellyn-Jones2011, Llewellyn-Jones2013) draw a sample list of criteria used to assess the court interpreter's presentation of self, participant alignment, and interaction management, which are summarised in Table 1.

As an example to which could illustrate Table 1, Llewellyn-Jones2014 report on the role-space that they adopted whilst interpreting during a court hearing. Their presentation of self was low as they introduced themselves as the interpreters, and then they were sworn in, but they did not provide the other participants with any further information about themselves. Their interaction management was quite high as they could seek clarification, and they could ask for

Table 1: Sample list of role-space criteria

Presentation of Self	Interaction Management	Participant Alignment
The interpreter...		
introduces herself/takes the oath or affirms	requests for clarification or repetition	addresses specific participants directly
refers to herself as “the interpreter”	manages turn-taking	provides feedback and back-channels
gives insights into her personal likes/dislikes	requests specific actions	explains some aspects of the interpreting process
answers direct questions	requests change in the environment	smiles when a participant makes a humorous contribution
divulges personal information about herself		reads body language/establishes eye-contact

questions to be reframed. However, they were more reluctant to regulate the interaction between participants. Finally, their participant alignment was limited to ensuring participants’ understanding of the proceedings. As a result, they aligned equally between the participants, but they felt that their alignment was very low. Their role-space model is represented in Figure 2.

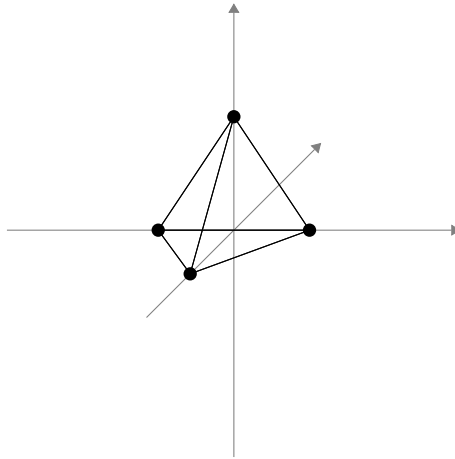


Figure 2: Llewellyn-Jones2014’s (Llewellyn-Jones2014) role space model based on their court experience described above

## 4 Methodology and research design

From a methodological viewpoint, this study was anchored within Actor-Network Theory (ANT). Such an approach enables researchers to examine human and non-human actors as forming part of an interactive network.<sup>3</sup> From an ANT ontological viewpoint, reality stems from the interplay created between human and non-human actors, and epistemologically, reality may be plural and unravelling it “requires identifying and following those actually involved in [the network’s] creation” (Bonner2013). Applying ANT to this study enabled the researcher to examine how the interpreter, the other court participants, and the VCI equipment interacted, through the eyes of the interpreter, to create a network during the court hearing within which the interpreter would express her role-space.<sup>4</sup>

After obtaining ethical approval from the University of Salford to carry out the doctoral project, 1,150 prospective participants who were all members of the National Register of Public Service Interpreters (NRPSI) were contacted by email in 2014. Thirty-nine expressed an interest and, in the end, eighteen practising court interpreters in England and Wales were interviewed. When applying Seidman2006’s (Seidman2006) principles of sufficiency and saturation of information, the number of participants was deemed sufficient.

In terms of participants’ profiles, there were sixteen women and two men sharing, between them, a wide range of language combinations (mainly European languages, but also Arabic, Chinese, and Turkish). All participants had a Diploma in Public Service Interpreting (DPSI), or equivalent, which at the time was a requirement to become a court interpreter. Several participants had more than one qualification, and they often combined a DPSI with a degree or an MA in languages or in Translation and Interpreting Studies. Most participants (twelve) had at least ten years’ experience in court interpreting. Their experience in interpreting in VCI was somehow more limited with most participants having interpreted on 10 or fewer occasions in VCI A and/or B.

In line with the ANT’s methodological stance, semi-structured interviews were conducted with the eighteen participants. The interview pointers were based on Table 1 in order to ensure that enough information was collected so that a role-space model could be drawn for each participant. The interviews were conducted either face-to-face or via Skype<sup>5</sup>. The recordings were then transcribed verbatim

---

<sup>3</sup>For more information on ANT, Latour2005 offers a good introduction on the interaction and networks created between humans and non-human entities.

<sup>4</sup>For more information on the extent to which ANT and role-space are compatible, see Devaux2017b.

<sup>5</sup>For a more in-depth discussion on conducting interviews in face-to-face or via Skype, see



and coded using NVivo. The corpus gathered accounted for 12.50 hours of recording.

## 5 Data analysis through role-space

The eighteen participants taking part in this study expressed different role perceptions, and the results are summarised in Tables 2, 3, 4, and 5.

Table 2: Role-space in vci A

	Presentation of self	Participant Alignment <sup>a</sup>	Interaction Management
P1	Very low	> Court	Very low
P3	Low	>Court	Low
P5	Low	Equal	High
P6	Low	Equal	Quite high
P7	Low	> Court	Low
P8	Very low	> Court	From low to high
P9	Low	> Court	High
P14	Low	Equal	From low to high
P15	Low	Equal	From very low to high
P16	Low	> Court	From low to quite high
P17	Low	> Court	Low
P18	Low	> Court	Low to quite high

<sup>a</sup>The sign “>” designates the side towards which the interpreter aligned.

The reasons explaining this assessment are summarised below. However, as some participants’ models differed greatly from those presented in Table 2 and Table 3, their models are discussed separately in §5.4. Due to word constraints and in light of the various role-spaces created, this section offers a brief summary of the data analysis, and a more in-depth analysis is provided in Devaux2017b.

### 5.1 Presentation of self

In vci A, P1 and P8 could not introduce themselves and/or were not sworn-in when all parties were in attendance. P1 stated that she could have been “the cleaner, (...) the woman with the microphone.” Similarly, P8 believed that it was

Devaux2017b.

Table 3: Role-space in vCI B

	Presentation of self	Participant Alignment	Interaction Management
P1	Very low	Equal	Low
P2	Low	Equal	Low to quite high
P3	Low	>Defendant	High
P4	Low	> Witness	Quite high
P5	Low	Equal	High
P6	Low	Equal	Quite high
P11	Low	> Witness	High
P15	Low	Equal	Low to high
P16	Low	> Witness	Low to quite high
P18	Low	Equal	Low to quite high

not obvious to the defendant that she was the interpreter. P1 added that she believed it was difficult for interpreters to be perceived as impartial as the defendant was on the other side of the screen. Similarly, in vCI B, P1 stated the introduction/sworn-in process was missing. As such the presentation of self for these participants in these settings was deemed very low.

The other participants reported that they had been sworn-in. They had introduced themselves, as the interpreter, to all the parties on both sides of the screen. Most also believed that impartiality was not impaired by the use of vCI equipment. However, they did not divulge any further information, for instance about themselves. As such, their presentation of self was deemed low.

## 5.2 Participant alignment

Four participants in vCI A and six in vCI B reported that they were able to align equally between the participants on both sides of the screen. Although some reported that their ability to hear/see well may be slightly affected by the use of vCI equipment, it had never been so poor that it had affected their interpreting performance. In fact, one participant (P15) even reported that it was easier to see/hear the court participants in vCI A than when interpreting face-to-face from the dock. The participants were also able to replicate body language, give feedback, and intervene to explain cultural references, when needed.

The other participants' experience differed greatly. In vCI A, they reported that they encountered technical difficulties that impacted on their ability to hear

and/or see the remote party well. They often stated that the sound was echoing, and P1 even compared the setting to a mausoleum. Furthermore, they felt that they could not read the witness or defendant's body language to obtain feedback or backchannel. Some also stated that the use of vCI equipment had a negative impact on the proceedings as they could not interpret all the content, cultural references, and/or build a rapport with the defendant in vCI A in order to adapt their terminology. Interestingly, the reason P9 aligned more towards the participants in court differs. She did not report on her inability to align with the defendant, but by sitting next to the judge, the close physical proximity made her over-align towards the participants in court. In vCI B, participants stated that it was more difficult to replicate body language as the screens were too small. They also reported that although it was easier to build a relationship with the defendant/witness, it was more difficult to reproduce this with the participants in court. As a result, these participants aligned more towards the party with whom they were physically present.

### 5.3 Interaction management

Some participants perceived that their interaction management had been either very low, low, quite high, or high. Their reasons differed and are summarised below.

P1 felt that her interaction management was very low in vCI A. She argued that the defendant was so removed from the process that no interaction could take place. Other participants' interaction management was low as they felt that the working environment was more daunting, and they were reluctant to interrupt the proceedings to ask for clarification. They also felt that the use of technology made the defendant less likely to interrupt the proceedings. On the other hand, some participants' interaction management was high or very high, be it in vCI A or vCI B. For instance, P6 and P9 said that it was easier to ask for repetitions and clarification as they were in clear view of the judge in vCI A. P9 also said that she always mentioned to the defendant that he could interrupt her at any time, if he did not understand. On several occasions, the defendant then interrupted her in vCI A, and she was able to manage the flow of interaction. Similarly, P4's interaction management in vCI B was quite high. Although she did not encounter any instances of over-lapping turns, she managed the interaction by asking the court to speak in "smaller chunks" as technology was used. Interestingly, she also stated that the defendant was less engaged in the proceedings, despite the interpreter being present in the same room. P5 managed the interaction by telling the defendant "Please don't talk, I am listening", and then she informed the court

that the defendant had a question. In the same vein as P4's approach, P5 required the court to speak one sentence at a time.

The above participants' interaction management could be described as static in the sense that they perceived it to be either very low, low, quite high, or high in VCI A/B. Other participants perceived that their interaction management could be expressed alongside a continuum ranging from very low to high. These participants felt that there was no need/hardly any need to intervene during the VCI court hearing, be it in VCI A and/or VCI B. Hence, their interaction management was very low/low. Nonetheless, had the need occurred, they stated that it would be possible to have a quite high/high interaction management, and that the use of VCI technology would not impair their abilities to manage the interaction. The reasons put forward by the participants as to why interaction management in VCI was very low/low differed, and they can be summarised as follows:

1. the hearings tend to be quite short, hence reducing the opportunity to encounter cultural references;
2. interpreter's expectation that the defendant would show enough respect to the court not to intervene;
3. over-lapping turns tend to be more frequent at police stations;
4. as the defendant appears remotely, he is less likely to intervene in court;
5. the interventions were "quite clear and straightforward" (P15).

It would not be feasible to create a role-space model for each of the participants in this article.<sup>6</sup> However, the general shape of their models could be divided into two categories: those who perceived their interaction management as static, therefore creating a four-face-pyramid model (Figure 3), and those who perceived their interaction management as a continuum, thus forming a five-face-pyramid one (Figure 4).

#### 5.4 Split role-space

Finally, some interpreters created one role-space for the participants in court and another one for defendant and witness. Their role-space axes are summarised in Tables 4 and 5.

---

<sup>6</sup>A detailed role-space analysis for each participant and their role-space model is available in Devaux2017b.

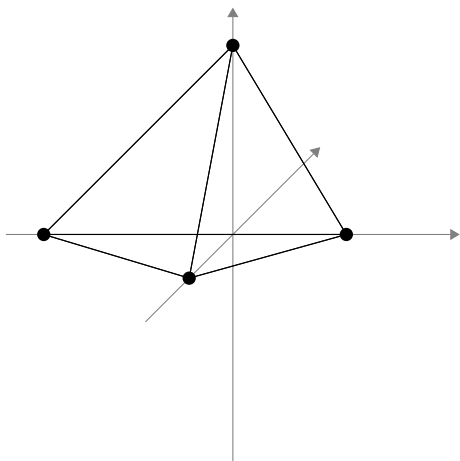


Figure 3: P 9's four-shape model in vCI A

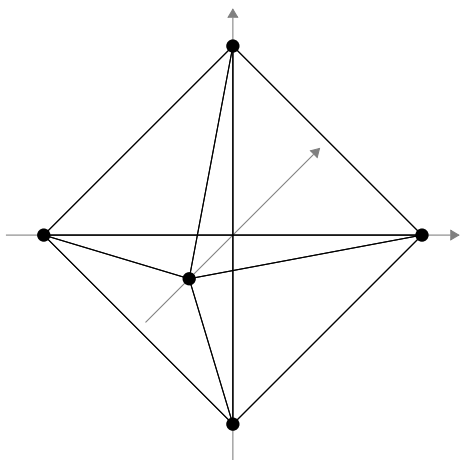


Figure 4: P15's five-face pyramid shape model in vCI A

Table 4: Split role-space in vCI A

	Presentation of self		Participant Alignment	Interaction management	
	Court	Defendant		Court	Defendant
P10	Low	Very low	> Court	High	Quite high
P12	Low	Very low	Equal	Low to quite high	

Table 5: Split role-space in VCI B

	Presentation of self		Participant Alignment	Interaction management	
	Court	Defendant		Court	Defendant
P13	Very low	Low	> Defendant and solicitor	Very low	Quite high
P14	Very low	Low	Equal	Low to quite high	
P17	Very low	Low	Equal	Low to high	

The reasons justifying the above participants' alignment are similar to those mentioned previously. Therefore, they will not be analysed again, and this subsection will focus on their presentation of self and/or interaction management.

In VCI A and VCI B their presentation of self was low with the co-located party, and very low with the remote participant(s). They could introduce themselves as the court interpreter with their co-located party. However, they could not replicate the process with the remote party, or even be sworn-in, for instance. Also, P12 raised some concerns on how she could be perceived as impartial by the defendant in VCI A as she was seen "on the same side" of the court, and P13 thought it was more difficult to establish an atmosphere of trust with the participants in court in VCI B. Hence, their presentation was low with the co-located party and very low with the remote participant(s).

P12 in VCI A, and P14 and P17 in VCI B perceived their interaction management alongside a continuum ranging from low to quite high/high. Again, the reasons were similar to those for the participants above, so they will not be discussed here.

The interaction management for P10 in VCI A and P13 in VCI B differed between the participants in court and the remote defendant/witness. P10 believed that the use of VCI equipment did not impact on his ability to manage the interaction with the participants in court. Yet, P10's interaction management with the other side was lower as "giving the conversation a rhythm [had been] much more difficult" (P10) with the defendant. Similarly, P13's interaction management was very low with the participants in court. Although he had encountered difficulties hearing parts of the court interventions, he did not intervene to notify the participants in court of the issue. Nevertheless, his interaction management with the solicitor and the defendant was quite high as they were sitting together, and he could attract their attention, and ask them for repetitions/clarification.

As a result, these participants' role-space model was split, and their shapes differed greatly from the other participants, as illustrated by Figures 5 (P10), 6

(P12), 7 (P13), 8 (P14), and 9 (P17). Worth noting is that P12 in vci A and P14 and P15 in vci B created two 3D role-space models. In order to preserve their model's readability, it was decided to split them into two graphs: one representing her role-space model with the participants in court (on the left hand-side), and one with the defendant (on the right-hand side).

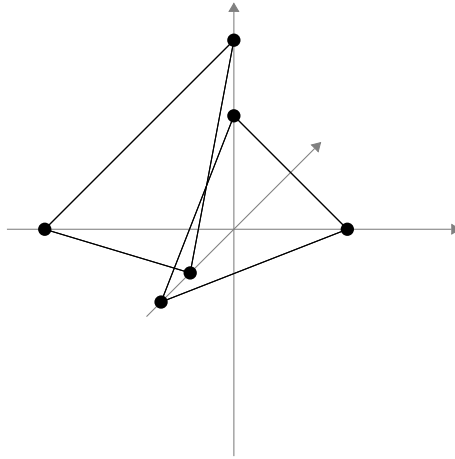


Figure 5: P10's role-space model in vci A

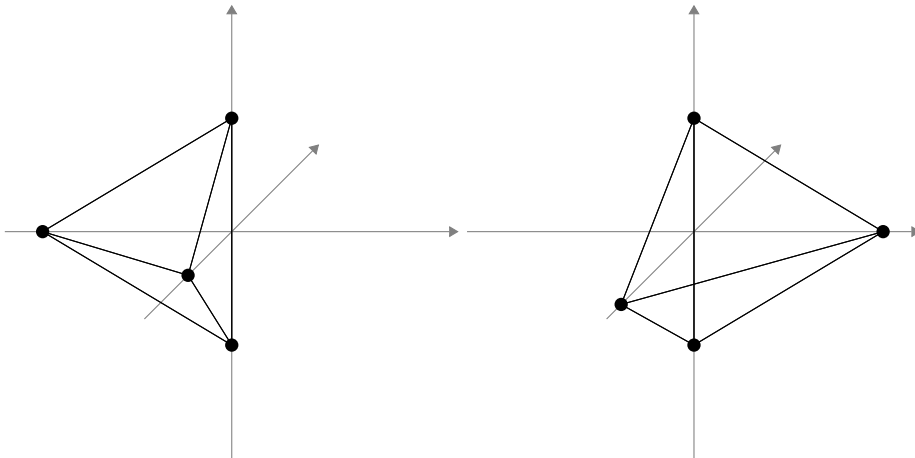


Figure 6: P12's role-space model with the court (left) and the defendant (right) in vci A

The participants perceived their role differently in vci A and/or B, and, in fact,

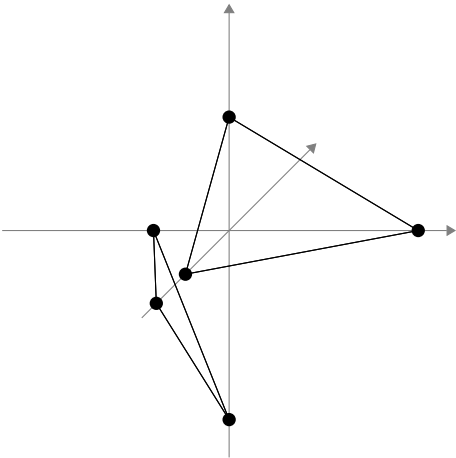


Figure 7: P13's role-space model in vci B

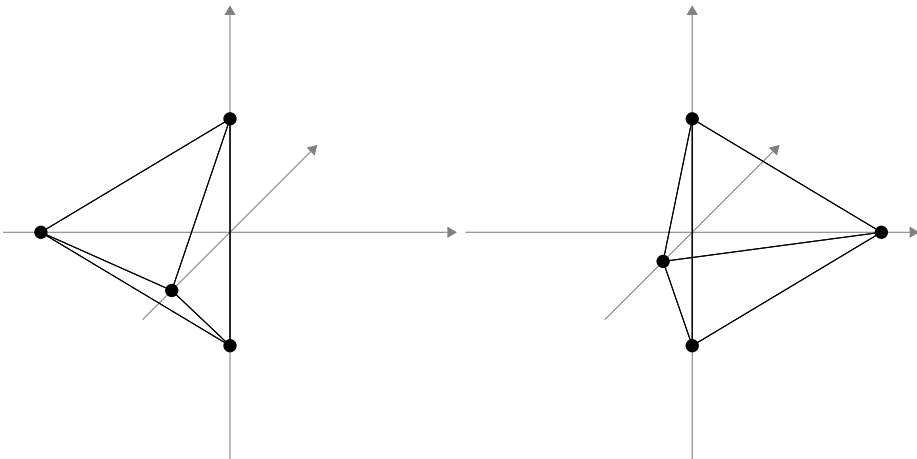


Figure 8: P14's role-space with the participants in court (left) and the witness (right) in vci B



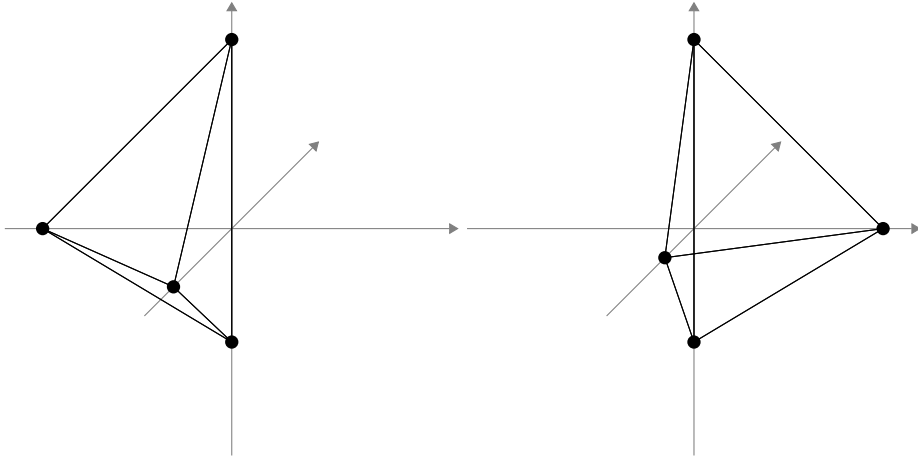


Figure 9: P17's role-space model with the participants in court (left) and the defendant (right) in vCI B

very few participants shared the exact same role-space (except, for instance, P3 and P7 in vCI A, and P12 in vCI A and P14 in vCI B). Although there were many different perceptions, their role-space can be grouped into three main categories: a fixed, a continuum, or a split role-space. Also, the use of equipment affected the participants' axes to various extent, and the distance between the participants meant that some interpreters were not always able to present themselves, manage aspects of the interaction, and/or align equally between the court participants. Worth noting is the fact that some participants mentioned that such equipment could also improve parts of the court proceedings. Notably, it was easier for some interpreters to seek clarification in vCI A.

## 6 Discussion of findings

Building on the above findings, this Section discusses the participants' perceptions of their role in light of the literature review, and it suggests some recommendations for training.

### 6.1 True-to-life experience

vanRotterdam2011 argue that the use of vCI equipment cannot represent a true-to-life experience due to various factors affecting the proceedings (such as potential poor quality of sound/picture, establishing eye contact, participants' re-

actions/interactions, etc.). In this study, parts of the data confirm that the absence of body language and back channelling, for instance, were highlighted as factors affecting some participants' experience, and these findings align with other studies which analyse the impact of the use of technologies. For instance, **Radburn-Remfry1994** argues that, in a mono-lingual setting, participants in court may feel more emotionally detached from the defendant during vci hearings. **Hodges2008** raises questions regarding the working relations between the defendant and the defence counsel. Supporting their studies, P1 felt that the defendant had not been taking part in his own hearing as he was too divorced from the proceedings. P1 believed that at this point, the right to see due legal process taking place could be questioned. This would support the idea that vci cannot replicate a true-to-life experience. It is worth noting that the participants' experience was not homogeneous, and it was even contradictory in some parts. For instance, P9 believed that when the defendant did not speak with a strong regional accent, there were no differences regarding whether the hearing was conducted in face-to-face or vci A mode. This plurality of interpreters' perceptions reflects this study's epistemological stance and the interpreters' multivocality. The array of perceptions is a window opening onto the actors' various realities, which suggests that a true-to-life experience can be a rather subjective notion, and from the interpreters' viewpoints vci equipment can impair or improve aspects of vc-conducted court hearing.

## **6.2 Factors affecting the interpreter's perception of her role**

The interpreters' different role perceptions in this study are not a new phenomenon. Some studies in is offer several factors justifying the interpreters' perceptions of their role differently in face-to-face settings, such as qualifications (**Martin2008**) and cultural acceptability (**Merlini2009**). One could also question the extent to which professional experience shape the interpreters' perceptions. All the participants in this study were DPSI qualified – some passed the same type of DPSI (e.g. law), and some were trained in the same centres. Yet, their role-spaces were different. Furthermore, some participants shared the same culture, but their role-space models differed. Finally, P1 and P2 in vci B have many years of experience as court interpreters (20 and 15 years, respectively), but their role-spaces also differ. Similarly, P6 and P10 have both interpreted over ten times in vci, but the role-spaces created are different. This tends to suggest that if qualification, cultural acceptability, and professional experience influence the participants' role perception in this study, these could only be factors partly contributing to shaping such perceptions. Nevertheless, the recurring denominator when analysing

the interviews seems to be the extent to which the participants felt that the use of vci equipment had limited parts of their role-space.

Due to the use of vci equipment, some participants had a very low presentation of self as they could not introduce themselves and/or be sworn-in. In her study, **Fowler2013** observed that some interpreters were not introduced at the start of the hearing. In such instances, there was “a tendency for the interpreter to defer to the court in matters which were properly part of their own professional remit” (**Fowler2013**). P1 raised the sitting arrangement as a potential issue with being perceived as impartial. This has been identified as a potential issue in the Avidicus 2’s research report, where their findings show that “the seating arrangements gave the impression that the participants on one side of the video link spoke ‘as one’ or ‘could be perceived as one’” (**Braun2013**). However, it is worth noting that other participants perceived that the seating arrangement was in fact improving aspects of their role-space as they were no longer interpreting from the dock at the back of the courtroom.

Some participants also aligned more towards one party as they felt that the technical difficulties encountered, and/or the lack of feedback or back-channelling opportunities, had not enabled them to establish a rapport with the court participant(s) on the other side of the screen. As a result, their participant alignment with the remote party had been lower. These findings align with **Rombouts2011**’s (**Rombouts2011**) and **Napier2011**’s (**Napier2011**) studies, which reveal that it is more difficult to establish a rapport with the remote party, or with (**Braun2016b**), who asserts that vci “entail[s] a reduction in the quality of the intersubjective relations between the participants.” This study also shows that when the interpreters’ participant alignment differs between actors, there had been a greater tendency to align towards the participants in court, rather than the witness or, to even a lesser extent, the defendant. Although the findings partly corroborate the difficulty to establish a rapport with the remote party, they also suggest that there are alignment disparities between the court actors and, amongst all the participants, the defendant is the party that the interpreter may be the least willing to over-align towards.

Finally, a few participants believed that the use of vci equipment had slightly reduced their interaction management, but overall it had remained quite high, which is similar to **Llewellyn-Jones2014**’s (**Llewellyn-Jones2014**) experience as court interpreters. Other participants had a high interaction management, whilst others had perceived it as ranging from low to high. Only a few participants had felt that their interaction management had been low. To some extent, the data gathered in this study concurs partially with the literature review, in the sense

that court interpreters have to manage the interaction by giving turns, for instance (Angelelli2003; Llewellyn-Jones2014; Martin2008), and they have to do so even more when technologies are used, to the extent that they become fully-fledged independent actors (Lee2007; Rosenberg2007). Similarly, Braun2016a argues that the discourse is more fragmented in vci-conducted legal proceedings, and it is therefore not surprising that many participants had a quite high or high interaction management. What remains unclear, though, is the reason why some court interpreters failed to intervene in order to re-balance their interaction management.

The use of vci equipment affected the participants' axes to varying degrees and very few participants' role-space was similar to Llewellyn-Jones2014's (Llewellyn-Jones2014) court experience (Figure 2). Despite such differences the shapes of most participants' role-space models were similar to the types discussed in Llewellyn-Jones2014's (Llewellyn-Jones2014) work. Indeed, some interpreters perceived their role as a 3D fixed entity, whilst others created a 3D continuum. However, unlike Llewellyn-Jones2014's (Llewellyn-Jones2014) models in face-to-face settings, this study reveals that the interpreter can adopt a third type of model, whereby she splits her role-space into two sub-spaces. In these instances, this study's participants felt that the use of technology had a limited impact on their role perceptions with their co-located party, but it restricted aspects of their role-space with the remote party.

### 6.3 Recommendations for training

As mentioned by some participants and as confirmed when examining the IoL2015's (IoL2015) *Handbook for Candidates* sitting the DPSI examination, it seems that vci training does not form an integral part of the DPSI curriculum. Therefore, given the fact that vci is used in court, and given the impact that the equipment can have on the interpreter's perception of her role, it is recommended that training in the use of vci in court be offered.

In order to ensure that prospective court interpreters possess more than a conceptual understanding of vci, DPSI centres should give students the opportunity to observe proceedings taking place in both vci A and B modes. It is also important that trainees are given the opportunity to practise role-plays in these two modes. Although centres may not be equipped with vci technologies meeting the TelecommunicationUnion2009's (TelecommunicationUnion2009) H323 Recommendation, trainees could nonetheless practise role-plays using cruder technologies such as Skype. Furthermore, the aim of this article was not to develop or assess a curriculum for trainee psis. However, centres could usefully develop resources based on the Braun2011c or Avidicus 3 training outline.

Participants also reported that court actors sometimes lack etiquette in terms of vci equipment and its use, making it more difficult for the interpreter to hear the proceedings. For instance, court staff or members of the public would leave the courtroom mid-hearing, court members would rustle papers near their microphone, or they would not speak in their microphone. Therefore, it is essential that legal practitioners also receive training in using vci equipment. Such training would need to make specific reference to conducting a bilingual vci hearing in the presence of an interpreter. As mentioned by some interpreters, training should not take place independently from the other participants, but rather all the court participants should also jointly train in using vci equipment. This confirms the recommendations put forward by other scholars such as **Braun2011a** and **Fowler2012**.

It is worth noting that such training may also be relevant to practising court interpreters. The NRPSI has more than 2,000 registrants who are dispersed over a large geographical zone. A means to ensure that they are trained in vci mode could be for the NRPSI to offer CPD sessions to its members on this interpreting mode.

## 7 Conclusions

This article reported on the findings arising from **Devaux2017b**'s (**Devaux2017b**) doctoral thesis. Based on the analysis of eighteen interviews conducted with court interpreters in England and Wales, it emerged that the interpreters perceived their role differently in vci, and they created different role-space models. The use of vci equipment affected various aspects of their presentation of self, participant alignment, and/or interaction management. Given the increasing use of vci equipment in court, this article also made some recommendations for training court interpreters. Although this study includes some limitations (e.g. a limited number of participants, who were all DPSI qualified), it paves the way to avenues for further research. To give a more generic picture of the court interpreter's role, this study's findings could be complemented by recruiting court interpreters without a DPSI (or equivalent). Another avenue to complement this study would be to observe court interpreters' role when vci is used, and to interview the other court participants so that the impact of the interpreter's role in vci could be examined. Furthermore, role-space is a relatively new theoretical framework and, as such, it would benefit from further empirical studies across various settings. Finally, it emerges from this study that some interpreters created a split-role model. The effect that this new model may have on the court

participants and on the overall interaction is unknown and it would deserve further exploration.

## Chapter 6

# Presence? Remote? Remotely present! New technological approaches to remote simultaneous conference interpreting

Klaus Ziegler

AIIC Technical Committee

Sebastiano Gigliobianco

SDI München

Since the 1970's there have been several approaches to test and implement remote interpreting as complementary interpreting modality in addition to the traditional and proven interpretation on site. The reasons for experimenting with remote interpretation in conference settings are manifold and can generally be classified by economic aspects, availability issues or organizational matters. In this paper we discuss the preliminary results of a pilot study aimed at exploring how the limitations of remote interpreting described by the literature could be overcome using new technological advances in Information and Communication Technology. We discuss challenges and technological solutions for remote simultaneous conference interpreting from an interdisciplinary perspective and sketch out what future workspace for conference interpreters might look like.

Please check order of footnotes and punctuation throughout this chapter

## 1 Introduction

Since the 1970's there have been several approaches to test and implement remote interpreting as a complementary interpreting modality in addition to the traditional and proven interpretation on site, with all the parties involved (speakers,



audience and interpreters) being present in the same room, thus communicating in a more or less face-to-face scenario. The reasons for experimenting with remote interpretation in conference settings are manifold and can generally be classified by economic aspects (e.g. reduced travel costs for interpreters and/or speakers and/or audience), availability issues (e.g. no local interpreters available for a specific language combination) or organizational matters (e.g. an interpreter team can be hired within a shorter period of time, room design does not allow for interpreting booths or booths are simply not wanted to be seen in the room).

Despite the, to a certain extent, encouraging results of studies, tests and experiments carried out throughout history at UNESCO 1976<sup>1</sup>, the United Nations (1978<sup>2</sup>, 1982<sup>3</sup>, 2001<sup>4</sup>), the European Union (1992<sup>5</sup>, 2001<sup>6</sup>, 2005<sup>7</sup>) using different technologies for the transmission of audio and video signals from and to interpreters, there have always been two main factors preventing the large scale implementation of remote conference interpretation: technological limitations (due to insufficient availability of bandwidth for the synchronized transmission of sound and image with the necessary quality when transmitting via Internet or telecommunication network, or very high costs when using satellite communication, either exclusively or in combination with terrestrial transmission technologies) and the more or less general refusal of the use of so called “new technologies” by conference interpreters. Apart from measurable physiological factors like fatigue and stress, leading to symptoms such as headaches and concentration problems, interpreters used to complain about the unease they were feeling because of not “being there” (Mouzourakis2006) and not having the possibility to get the right feeling for the situation, attributing these psychological symptoms mainly to the limited view of the speaker and the audience as well as not having the possibility to interact directly with the other participants of the event.

In the last few years, general conditions for conference interpreting have been changing constantly due to globalization and altered market needs, on the one hand, but also due to digitalization and extremely fast developing information and communication technologies. The availability of hardware and software for dynamic monitoring and controlling of important parameters like lip synchronization, latency, video resolution and frequency response as well as network

---

<sup>1</sup>Kurz2000

<sup>2</sup>Chernov2004

<sup>3</sup>UNESCO1987

<sup>4</sup>Mouzourakis2006

<sup>5</sup>Kurz2000

<sup>6</sup>Europarl2001

<sup>7</sup>Roziner2010



infrastructures that allow for simultaneous transmission of high definition video and high quality audio signals via Internet, combined with latest video, virtual reality and augmented reality technologies might offer possibilities to overcome existing technological, physiological and psychological problems.

## 2 Interdisciplinary and terminological challenges

One major challenge when discussing “remote interpreting” as a method for the delivery of interpreting services resides in the fact that there are a lot of different concepts being used in practice by the different stakeholders when referring to this method. While this linguistic phenomenon can be explained, to a certain extent, by the fact that interpreters, in general, simply are not experts in the technical field and therefore use technical concepts without knowing exactly what the technical background for certain scenarios is and what the implications of a certain technical setup exactly are, we can observe that there is still no harmonized terminology being used even by technical experts and researchers.

Technically speaking, a video conference and video conferencing can be defined as

“a live, visual connection between two or more people residing in separate locations for the purpose of communication. At its simplest, video conferencing provides transmission of static images and text between two locations. At its most sophisticated, it provides transmission of full-motion video images and high-quality audio between multiple locations” (TechTarget2017)■

This definition shows that the same concept is being used for a huge variety of different technical setups.

When it comes to including interpreters in communication to facilitate the necessary translation between the languages that are being spoken by the participants in a communicative event, things become even more diffuse. The term ‘remote interpreting’ as one of the most widely used concepts, covers a whole range of technologically very different setups. These setups range from a traditional presence-based scenario where interpreters, main speakers and the audience are concentrated at one event location, and one or several secondary speakers are connected from a distance for a limited sequence of time, to a situation where none of the actors within the triad speaker-listener-interpreter is at the same location as the others.

Braun takes up this terminological challenge saying that

“Two main uses of telephone and videoconference communication can be distinguished in connection with interpreting. One of these, remote interpreting (RI), refers to the use of communication technologies to gain access to an interpreter in another room, building, town, city or country. In this setting, a telephone line or videoconference link is used to connect the interpreter to the primary participants, who are together at one site” (Braun2015),

thus excluding a setup where the participants are located at different locations. For this case, she introduces the concept of “teleconference interpreting to cover both telephone and videoconference communication” (Braun2015). Further on, she introduces as a separate term

“teleconference interpreting to cover both telephone and videoconference communication” (ibid.)

For disambiguity purposes, she then introduces the terms “telephone-based interpreting” and “videoconference-based interpreting” (ibid.), but adds that there are a lot of additional concepts being used in practice.

For the purposes of this article, we broaden up the perspective and adhere to the definition given in the recently published **ISO20108**, introducing the term of ‘distance interpreting’ (with ‘remote interpreting’ as admitted term), giving the definition of ‘interpreting of a speaker in a different location from that of the interpreter, enabled by information and communications technology (ICT)’.

Analyzing technological, communicative, cognitive, physiological and psychological aspects it becomes clear very soon that every single dislocation of one of the emitting or receiving elements (speaker, listener or interpreter) to a different location (thus becoming a distant location) has a considerable impact on the technological setup and the components and transmission channels needed to enable communication between the different parties involved.

Furthermore, communicative aspects like verbal and non-verbal communication are altered by the rearrangement of the setting, as elements like gestures, facial expressions etc. cannot be perceived directly anymore, and have to be captured, transmitted and reproduced again at the distant site in order to be made available.

As far as cognition is affected, the cognitive load generated by additional or, at least, altered receptive and productive tasks that interpreters will have to manage while performing, related to the sub-processes of listening to a source text, processing its content and re-producing that content in the target language, is also

being influenced significantly by every alteration of the setting (Moser-Mercer2005)■ Research has revealed that cognition is intimately linked to physiological processes that take place in the human body, especially the brain as the controlling unit, and that a variation of acoustic and visual input to interpreters has an impact on vital functional systems like the respiratory system, metabolism and others, leading to stress, earlier fatigue and other phenomena (Moser-Mercer2003).

Last, but not least, there is also a wide range of psychological factors that have to be considered when approaching “distance interpreting” from an interdisciplinary point of view. In an ideal setting, communication taking place with all participants being at the same location, in the same room and with no obstacles impeding direct mutual perception, allows the participants in the communicative event to make use of at least four out of the five basic human senses: touch, smell, see, hear. Whereas most of the related parameters, such as neuronal stimulation for haptic feedback, composition of ambient air for smelling, audio frequencies for hearing and light waves for vision can easily be measured, quantified and evaluated, other phenomena such as energy fields (sometimes referred to as mental, astral/emotional and etheric/physical bodies) that affect the balance of both the physical body and the non-physical mind, are much more difficult to capture and evaluate, although they might have an influence on human interaction and, by that, on communication.

Whereas many interesting aspects related to distance interpreting have been addressed based on other interpreting specializations, such as legal interpreting (see AVIDICUS projects 1–3), there have not yet been similar large-scale projects directly or indirectly related to distance simultaneous conference interpreting. However, several aspects have been studied on a smaller scale, such as stress and performance in remote interpreting (Moser-Mercer2003; Roziner2010), perception of remote interpreting by interpreters (Mouzourakis2006) or visual input (Rennert2008; Luisetto2016).

### 3 Solutions in the past

#### 3.1 Tests and experiments UNESCO, UN

Since the seventies, due to a quick development in telecommunication technology, big international institutions such as the United Nations or the European Union have begun testing new interpreting solutions to reduce the cost of conferences (UNESCO1987).

One of the first experiments with remote interpreting took place in 1976 when

the UNESCO organized its general assembly in Nairobi. The interpreters, however, were asked to work from Paris, which was connected to the capital city of Kenya through a satellite connection which provided an audio and video connection quality equal to that of a standard tv broadcast (Mouzourakis1996). The interpreters were not satisfied with their performance and stated that they were more tired and stressed than usual (Kurz2000).

A second experiment with remote interpreting was organized in 1978 by the United Nations. The conference was held in Buenos Aires and there were interpreters working both in Buenos Aires and in New York, the latter received audio and video signals through a satellite connection. The results showed that it was possible for the interpreters working in New York to achieve a high quality interpretation, (UNESCO1987) although the interpreters working in Buenos Aires were able to deliver a better performance due to their more intensive preparation and knowledge about the conference (Chernov2004).

Another experiment with satellite connection took place in Vienna in 1982 during the United Nation conference on the Exploration and Peaceful Use of Outer Space. At this conference, the interpreters worked in Vienna as well, but not in the same building (Andres2009). The communication was a success, but the interpreters complained of increased stress (UNESCO1987).

Although these experiments with satellite connection between the 70's and the 80's demonstrated that remote interpreting was possible, the huge costs and the increased level of stress led to the conclusion that this technology was too expensive and immature to be used yet (ibid. 26).

Thanks to the ongoing advancement of ICT and the growth of the European Union, new possibilities for remote interpreting arose. From the beginning of the nineties onward the European Union was the main driving force behind this development with the aim to create a central hub where the interpreters could work (Braun2011b).

One of the first positive results was achieved in 1999 in Vienna using ISDN technology during a UN Inter-Agency Meeting on Language Arrangements, Documentations and Publications which was held in Geneva and interpreted from Vienna. During this experiment, the transmitted audio signals were based for the first time on a frequency going up to 7 kHz. The conference hall in Geneva was recorded by three cameras and the pictures were projected in Vienna using a projection screen with a transmission rate of 384 kbps (Mouzourakis2006). The audience was satisfied with the performance of the interpreters and for the first time the interpreters were pleased with the quality of the audio transmission, although they criticized the video quality (Andres2009). Two months later

a second experiment was organized together with the International Telecommunication Union (ITU) and the École de Traduction et d'Interprétation (ETI). For this experiment two French booths were installed: one in the conference hall and a second one in a remote location. In this case the audio signal was encoded in MP3 format and the video signal was transmitted with a rate of 382 kbps (Mouzourakis2006). The audience alternately received the signal from the local and remote booth and was happy with the results. The interpreters, on the other hand, perceived the physical distance from the conference hall and had a feeling of losing control over the situation. Interestingly, saliva samples were taken from both booths before and after the conference, but contrary to what the interpreters stated, no relevant difference was noted in stress hormones (Moser-Mercer2003).

A new experiment was organized in 2001 in New York by the United Nations in which both ISDN and satellite connection were tested. The conference was recorded using three cameras and a 42-inch plus a 25-inch plasma screens were installed in front of the booths. With this experiment, the UN laid down the following minimum requirements for remote interpreting:

14 kHz sound (requiring 128 kbps) for sending floor sound to the booths (14 kHz) and 10 kHz sound (at 64 kbps) for sending interpretation back to the floor (10 kHz); 512 kbps for the image of the speaker plus 384 kbps for the floor/podium image. (Mouzourakis2006)

Two new studies took place in 2001 and 2005 at the European Parliament to test a connection through optic fibre (ibid. 64). The results of the first experiment showed that the interpreters were satisfied with the audio and video quality. However, they criticised the selection of pictures and the fact that they did not have a comprehensive view of the conference room (Europarl2001). Moreover, they stated that they felt uncomfortable and that the remote interpretation setting is overall more tiring (ibid. 22–23). The second experiment, which lasted five weeks, brought similar results: the interpreters were pleased with the audio and video quality although a complete view of the audience was lacking. Moreover, they felt alienated and isolated from the conference and the screens caused eye-burning, headaches, loss in concentration and higher tension and fatigue. They also stated that they felt their performance was of inferior quality while working remotely. Medical examinations, however, found no evidence of increased stress and a performance evaluation confirmed that the interpreters' performance working remotely was slightly inferior, but not enough to reach statistical relevance (Roziner2010).

## **4 Solutions at present**

### **4.1 LAN based solutions**

An example of a remote interpreting solution which is based on a local network is currently being used by the Directorate General for Interpretation of the European Commission (DG SCIC). In this setting, interpreters work from a conventional permanent booth located in another room at the same location, although occasionally mobile booths were also used, in front of which 4 high resolution screens are placed. Through the screens, interpreters receive the following images:

- On a 50" full HD screen with split view, an overview of the entire meeting table from two different angles is shown. Thanks to the big screen and the high resolution, the interpreters are able to see all the participants' faces.
- An image of the active speaker is transmitted on two 22" full HD screens which are placed laterally to the 50" screen
- A static shot of the presidency is displayed on a third 22" Full HD screen which is placed on top of the 50" screen.

In order to transmit the images from the meeting room to the interpreting room, optic fibre cables are used (Technical2016).

### **4.2 Solutions with external data transmission**

With the growing demand for a more flexible and less money consuming delivery of interpreting services in a globalized world, several attempts were made in the last decade to overcome restrictions and to make use of technologies allowing to leave the traditional presence-based scenario with a direct wired connection of all components of the conference system (see §3). For the purposes of this article and from a merely technological point of view, these technologies can be divided into solutions including the transmission of either only audio signals, or the transmission of both audio and video signals to the interpreters.

#### **4.2.1 Audio conferencing solutions**

One approach to overcome existing limitations of bandwidth and thus the impossibility for transmitting sound and image to the interpreters with the required

quality (see §3) consists in reducing the transmitted content exclusively to the audible signals. The most accessible and therefore most frequently used technology for this transmission relies on the use of telephony. In technical terms, a first very rough distinction has to be made between traditional land lines and mobile telephony, as the audible frequencies are mainly being transmitted either via wired connection (land line) or as waves through the air (although mobile telephony also includes sections where the transmission is being carried out by wire, as it is the case in the mobile devices themselves or the re-transmitting stations in between). The use of copper cable and the non-existence of fibre optic connections in certain areas, but also the necessity of handling billions of those connections at a time, still lead to a considerable cutting down of the audio frequencies being transmitted to a maximum of 8 kHz, even if land lines today are more and more often based on digital VoIP technology<sup>8</sup>. Whereas for consecutive interpreting the range of up to 8.000 Hz might be enough (often named ‘wideband audio’) because of the clear separation of the reception (hearing) and producing (talking) processes on the interpreter’s side, simultaneous interpreting with listening and talking at the same time is considerably affected by the loss of frequencies due to the masking effect generated by the interpreter’s own voice, making incoming frequencies on the same level not audible anymore (Jumpelt1984). As a result, the physical and cognitive effort for hearing will increase and, depending on the frequency of the incoming voice of the speaker and the interpreter’s own voice (note that e.g. female and male voices are located in different frequency ranges), the interpreter’s output quality will necessarily decrease despite all efforts to compensate for the missing input by interpreting strategies like variation of the voice-to-ear-span or additional pausing. It should not be forgotten that the interpreter has no possibility to compensate for the absence of audible input by visual input when using audio-only solutions. Even if these technologies have improved a lot and specially designed audio bridges allow controlling incoming and outgoing signals with a high level of quality of service even in multilingual scenarios, the above described acoustic parameters cannot be neglected.

#### 4.2.2 Video conferencing solutions

As far as solutions able to transmit sound and image at the same time are concerned, one of the most frequently used technology is based on video conferencing solutions. These solutions rely on the principle of separately capturing image and sound at the source with camera and microphone, encoding the data

---

<sup>8</sup>For a comparison of VoIP coding algorithms, see Singh2016

with a certain algorithm, transmitting the data packages using one or more parallel lines and decoding the packages again at the destination, making sound and image audible and visible again. These solutions necessarily depend on the availability of adequate and compatible equipment at all source and end points of the communication and the use of a common standard for encoding and decoding of signals (codec). One of the critical points with these video conferencing solutions in terms of quality simultaneous interpreting is the fact that standard codecs use a compressed file format for transmission of audio signals and compression is carried out either by cutting out certain frequencies, or by other procedures causing similar effects as described before for the audio only solutions<sup>9</sup>.

## 5 Technical requirements for remote interpreting

As described above, remote interpreting is a specific method of (conference) interpreting and covers a variety of scenarios of interpreting of a speaker at a different location from that of the interpreter, enabled by information and communication technology. The main aspect in these scenarios is that the interpreter is not physically present in the same room as all the other actors in the communicative event, thus not having a direct view of either the conference room and/or the speaker and/or the audience. To enable interpreters to do their job and deliver a quality service, some technical requirements need to be fulfilled. As we have seen before, different experiments have been carried out in order to identify the minimum technical requirements which are to be observed to assure performance quality.

One of the first attempts to define minimal requirements for remote interpreting was the *Code for the use of new technologies in conference interpreting* which was published by the International Association of Conference Interpreters (AIIC) in 2000 together with other national interpreter's association, the European Parliament, the European Court of Justice and the joint conference and interpreting service of the European Union and the World Customs Organisation (Korak2010). Without going too much into detail, AIIC states that interpreters' work conditions in remote settings must comply with the requirements set out in ISO standards 2603 and 4043 (1998 editions, withdrawn in 2016 and replaced by a new set of ISO standards 2603, 4043, 20108 and 20109) which define the work environment of interpreters in mobile and permanent booths. Moreover, the following requirements are to be fulfilled:

---

<sup>9</sup>For more information on techniques for video compression see Wiegand2003.



- All frequencies between 125 and 12500 Hz are to be transmitted.
- Interpreters must receive high definition images of the speaker and other participants.
- The interpreter shall work no more than two hours per day (AIIC2000).

A more comprehensive study commissioned by the European Commission was conducted in 2010 by the *Fraunhofer Institut* to evaluate the minimal requirements of video and audio quality for simultaneous interpretation<sup>10</sup>. During this study, conference interpreters were asked to evaluate different audio and video signals to assess the impact of transmission quality on their performance. The result of the study is a guideline which states that:

- All frequencies between 125 and 12500 Hz must be transmitted, although frequencies starting at 75 Hz should also be included in the range
- Video quality should be at least 1280×720 at 50 Hz with a ratio of 16:9
- Audio must be synchronized (lip synchronisation) with the video track with a maximum value of -25 ms or +95 ms

The work of the *Fraunhofer Institut* was used as a starting point to draft the new ISO-Standards 20109 “*Simultaneous interpreting – Equipment – Requirements*” and 20108 “*Simultaneous interpreting – Quality and transmission of sound and image input – Requirements*”, which not only raised the minimum requirements, but also added new ones concerning sound and image transmission:

- All frequencies between 125 and 15000 Hz +/-3 dB must be transmitted (ISO20109).
- Image quality must be good enough to avoid blurring and freezing of the video.<sup>11</sup>

---

<sup>10</sup>This study was originally published in 2010 by Sporer, Thomas; Fischer, Jens-Oliver; Liebetrau, Judith; Fröhlich, Daniel; Schneider, Sebastian; Kämpf, Sven under the title “Definition of an Objective evaluation method for assessing the minimal quality of digital video and audio sources required to provide simultaneous interpretation. Final Report. Fraunhofer IDMT.” When this article was drafted, it could be found at [http://ec.europa.eu/dgs/scic/docs/working\\_with\\_int/inf\\_tech\\_group/2010-12/fraunhofer\\_study\\_2010\\_final\\_report.pdf](http://ec.europa.eu/dgs/scic/docs/working_with_int/inf_tech_group/2010-12/fraunhofer_study_2010_final_report.pdf)

<sup>11</sup>The draft of the ISO-standard 20108 contained more detailed specifications regarding video and transmission requirements: video quality must be at least 720p at 50 Hz or 1080p at 25 Hz and the signal must be compressed using at least H.264 at 1152 kbps. Moreover, the packet loss value should not exceed 0.2%, jitter should be lower than 15 ms and the latency (roundtrip) in the system shall not exceed 200 ms.

- Audio must be synchronised with the images with a maximum delay of 45 ms or advance of 125 ms.
- Latency (from the source to the interpreters) must be lower than 500 ms (ISO20108).

## **6 Future workspace**

Based on the technical parameters exposed in §5, the authors started an experimental research project with the objective of modelling a future workspace for conference interpreters while performing remotely in simultaneous mode. For these purposes, the framework that was chosen for the experimental study was based on a standard working environment for conference interpreters, including, amongst others, a soundproof simultaneous interpreting booth and a hardware interpreter's console as audio/video interface. Apart of these technical aspects, the experimental setting was also characterized by some specific markers in terms of interactivity and communication patterns, generally assigned to conference interpreting in literature, such as a mostly monological discourse pattern, interpreting into one language, no possibility for interaction with the speaker during his intervention, a symmetric communicative setting with speakers on the same educational level, the same linguistic code used by both speakers, and no variation of speaker's registers (Angelelli2000). All tested scenarios in the experiment were designed with a view to possible implementation in conference interpreting "hubs" with a basic technical setup similar to the one that can usually be found at conference venues.

### **6.1 The Experiment**

The experiment took place in Düsseldorf in collaboration with pcs Professional Conference System GmbH. The aim of this experiment was to test three different remote interpreting settings:

- using a 65-inch screen with a picture-in-picture function
- using a camera remotely controlled by the interpreter, and
- using a 360-degree camera and Virtual-Reality-Glasses.

Common to all three parts of the experiment was the text which the interpreters had to interpret in simultaneous mode. Two speakers were asked to speak

for a total length of circa 10 minutes per scenario. The dialogue was not prepared but rather improvised and the only rule the speakers had was that the first five minutes of the speech had to be informal and the last five of a more technical nature. Since according to the effort model of Gile<sup>2009</sup> interpreters need to distribute their concentration among the different tasks of simultaneous interpreting, the different degree of difficulty of the speeches aimed to test whether interpreters were still able to operate the new device in the booth under different degrees of stress. The interpreters were not informed about the nature of the text they would have to interpret, for the purpose of eliminating preparation as a variable from the equation and to assure that they had to concentrate on the interpreting effort as well. Moreover, since the interpreters had a direct control on the video signals being transmitted, two speakers were intentionally selected in order to force the interpreters to adjust the video settings based on which speaker was talking in order to assure that the new device was actively used during the experiment. The dialogue was held in German and each of the two interpreters was working from his A into his B language. Since the length of each dialogue was only ten minutes and pauses between the scenarios were planned, each interpreter worked alone in the booth while the other was waiting in a separate room, to guarantee equal conditions. Moreover, having each interpreter working alone assured that they were forced to operate the extra technological feature while interpreting, to determine whether and to what extent this would have an impact on their performance. After the experiment, the interpreters were asked to fill out a questionnaire to evaluate the different settings and technological setups. The questionnaire consisted of 28 questions divided into two parts: first a more generic part to create a profile of the subjects with questions about their age, experience, linguistic combination and whether or not they already had experience with remote interpreting, and then a second more specific one, in which, for each scenario, the interpreters were asked to summarize their personal opinion about these new solutions. For example, they were asked whether they could operate the extra devices in the booth, whether they felt they had to put extra effort into it and whether these new solutions were, according to them, better or poorer than a standard video and audio transmission which they could not control. Moreover, for each scenario, they were asked to evaluate the quality of the video transmission on a scale from 1 (poorest) to 5 (best), to share any personal opinion, comment, or criticism and finally to state whether or not they would like to see this solution implemented in the future.

The subjects in this study were both professional conference interpreters holding a degree in conference interpreting. Both started their career more than 16

years ago and stated having worked as conference interpreters between 51 and 100 days per year, and can therefore be qualified as ‘experienced’ professional interpreters. In terms of internal validity of the study it is also important to mention that they stated having had very little or no experience at all with remote interpreting in conference settings. Due to the very small sample of this study, the results of this experiment cannot be generalized. The study aimed primarily at testing the feasibility of these new technologies. To achieve external validity, the study would have to be reproduced on a larger scale and conducted with an appropriate methodological approach.

### **6.1.1 Picture-in-Picture**

The picture-in-picture solution was the first scenario being tested. For this part of the experiment, the speakers were in the conference room, in which two different cameras, each pointed at one speaker, were recording the event. The audio signal was recorded with a wireless microphone and fed together with the video signal to the Extron SMP 351 recording and streaming processor, which streamed the audio and video tracks over the local network of PCs to the interpreting booth. To guarantee lip synchronisation the audio signal was delayed by 275 ms before being fed to the Extron. In another room, a 65” monitor was placed 95 cm from the booth, guaranteeing roughly a distance of 155 cm between the screen and the eyes of the interpreters. According to Causo2011’s guidelines (Causo2011) the booth was placed in such manner that the windows’ frames of the booth would not obstruct the view of the interpreters. The video signal from the conference room was shown to the interpreters using a picture-in-picture function meaning that both signals were transmitted simultaneously and that one image occupied a larger portion of the screen while the second video feed was displayed in a smaller format in the bottom right corner and the interpreters were able to switch between the two. According to the draft of the ISO standard 20108<sup>12</sup> (ISO20108), the video signal was reproduced with a resolution of 1920×1080 pixels at 30 frames per seconds.

The results showed that both participants reacted positively to this solution and they found it to be better than a static video signal, upon which they have no influence. None of the interpreters noticed any increase in workload due to the more technical nature of the second half of the text and stated that they had no problem in using the new device in the booth to switch between the two pictures.

---

<sup>12</sup>When the experiment was conducted, a final version of the ISO-Norm 20108 had not been published yet. The draft was therefore used for reference.

Both interpreters stated that they would like to see this solution implemented in the future.

### 6.1.2 Remote controlled camera

During the second part of the experiment, the speakers were recorded by only one camera placed in the middle of the conference table. From within the booth the interpreters were able to remotely move the camera 270 degrees horizontally and 90 degrees vertically on its axes. The camera was directly connected to the internal local network of PCs and the interpreters were able to move it using a mouse connected to a laptop in the booth. The audio signal was captured separately using a wireless microphone and delayed by 275 ms before being transferred to the interpreter's booth using a cable. With a resolution of 1080p by 25 fps, this camera respected the indications of the ISO standard 20108 as well.

The results of the questionnaires showed that this solution was also welcomed by the interpreters and that they quickly got used to the control of the camera via mouse and stated that this could rapidly become an automatism. This solution was also considered to be better than a non-controllable video transmission. None of the interpreters noticed the difference between the first and second part of the dialogue and had no problem with the camera control.

### 6.1.3 Virtual reality glasses

For the last part of the experiment, the interpreters still worked in a normal interpreting booth wearing VR-glasses of the company Elegiant which use a smartphone (in this case an Apple iPhone SE) to reproduce images. The smartphone was directly connected through Wi-Fi to the 360-degree camera used to capture the dialogue between the two speakers. Also in this case, the audio was captured with a wireless microphone and fed to the interpreting booth using a cable after being delayed by 500 ms. In order to fully exploit the potential of the camera and the VR glasses, the speakers were standing in front of each other in a big hall and asked to move around while talking as well as making use of gestures and mimics.

The opinions of the interpreters on this solution were very different from each other. The first interpreter found the VR-glasses an interesting alternative to a monitor although not very comfortable to wear for a longer period. The subject also stated that, after a short time, one gets used to blindly operating the console in the booth. The second interpreter, on the other hand, was not happy with this solution, finding it tiring for the eyes -- it must be noted that the second

interpreter wore glasses and because of the design of the VR-glasses it was not possible to wear both correction and VR glasses – and complained that it was very difficult if not impossible to operate the console in the booth or use a personal computer for terminology research.

At this point the authors need to clarify that especially this last part of the experiment was of extreme experimental nature and that the results are not particularly conclusive, also due to the small number of participants. The main goal of this experiment, however, was not to test whether the selected model of VR glasses is ready for a practical application in interpreting booths, but rather to make a first test on operability of VR glasses combined with traditional equipment for simultaneous interpreting in soundproof booths. The results of this test shall be used for additional and more elaborate studies in the future.

## **6.2 From virtual reality to augmented reality solutions**

The experimental study described in §6.1 suggests that the lack of direct view of the speaker during an event might be compensated for, at least to a certain extent, by the use of technologies that allow the interpreter to control the video input he feels he needs to process non-verbal elements of the content produced by the speaker himself. Nevertheless, a one-dimensional screen reproducing one or several images will always reduce the possibilities of perception of the setting captured with cameras, as there is still a very clear separation of the interpreter in his remote environment and the room where the original event is taking place. The feeling of separation from the action, often referred to as the feeling of not “being there”, might still be big enough as to prevent the interpreter from overcoming this psychologically relevant issue.

The last step in the experiment described above was to dive into a virtual reality scenario, making the interpreter feel immersed in the situation, e.g. making similar movements with his eyes and/or his head and body as he would be doing if seated in the room with the speaker and the audience, seeking for the necessary visual information to complete the audible content and render the entire intended message to the listeners. Whereas the use of a screen with a double video feed, controlled by the interpreter and reproduced as a picture-in-picture image on a wide screen was accepted by both subjects in the experiment and the additional control task didn’t seem to have any negative effect in terms of cognitive overload, the use of virtual reality glasses clearly showed that the physical separation from the real world and the traditional equipment placed in it (console for controlling audio input and output, and laptop for document and knowledge management) raises several cognitive and ergonomic issues, although the much

more dynamic and self-controlled setting was considered positive as such.

Following this line, the combination of both real and virtual elements in an augmented or mixed reality scenario, where computer-generated images are superimposed on user's view of the real world, would be the next logical step towards a practical solution for the challenges that remote simultaneous interpreting imposes. AR glasses already being offered on the market like Microsoft's HoloLens, for example, allow projecting images and possibly other virtual images into the vision field of the interpreter, while he can still see and control the real hardware components he needs for the delivery of the interpreting service. Any software application used for document and knowledge management could be moved from the real world into the virtual world, projecting only the image of the respective interface into the field of vision and allowing him to virtually manipulate the application with his hands. For these purposes, ergonomic aspects such as weight and wearing comfort of the device, cognitive aspects such as real or perceived additional workload and, of course, compliance with the technical parameters as set out for distance interpreting, as well as processing capacity of the processing unit will have to be studied in terms of usability and feasibility.

### 6.3 Interpreting hub solution

Considering that workspace for conference interpreters working remotely in simultaneous mode should take into account the relevant parameters as stated in existing ISO standards 2603/4043, as far as sound insulation, ventilation and ergonomics are concerned, ISO 20109, as far as equipment for simultaneous interpreting is concerned, and ISO 20108, as far as quality and transmission of sound and image input to interpreters is concerned, workspace solutions for distance interpreting should be designed accordingly, assuring that the most important parameters that allow for quality simultaneous interpreting are met.

Even if for economic reasons a solution with the interpreter working from his home office might seem the most obvious and easiest solution, there are a couple of issues that need to be considered when envisaging such solutions: First of all, in many regions of the world a dedicated internet connection with assured availability of the necessary bandwidth to transmit high definition images together with ISO compliant sound to the interpreter, is either non-existent or hardly affordable for a single interpreter. In addition, home office rooms would have to be equipped either with an appropriate interpreting booth or with components assuring compliance with the main parameters of ISO 2603/4043 in terms of sound insulation, ventilation etc.

A lack of possibilities for dynamic control of the internet connection while

interpreting, as well as data protection and confidentiality issues would also have to be resolved if home office workspace were to be used.

Particularly for multilingual events with more than two languages being spoken and interpreted at the same time, a promising approach in terms of Quality of Service both from a technical and an interpreting performance view are solutions with interpreting being performed in specially designed hubs, where interpreters would find all the necessary working conditions to deliver quality interpreting. Traditional ISO compliant booths as a minimum standard, or a specifically designed workspace ensuring the basic requirements, equipped with state-of-the-art digital interpreting equipment, dedicated internet connections with permanently available bandwidth according to the number of video and audio feeds needed, controlled ambient conditions with active regulation of air supply and carbon dioxide levels, immersive 3D environments with high resolution projection screens and interpreters being able to choose the desired visual input out of several video feeds, AR components, and technical support available on site would set the appropriate technical framework. In addition, offices with access to online and offline information sources for preparation before and during the event, rest rooms for relaxation and lounge areas for the necessary professional and social exchange within the interpreting team, would add other useful elements to overcome the alienation perceived by interpreters working in remote scenarios and environments that have not been developed specifically for this purpose.

Apart from these more technical aspects, any kind of hub solution would also have to address psychological factors such as measurable or perceived stress. Moser-Mercer<sup>2005</sup> states that “it appears that (...) interpreters seem to be under increased psychological stress when working away from the conference room, mostly because they experience a lack of control of the situation”. Even if there are no studies available yet, that would explain exactly which kind of control interpreters are missing<sup>13</sup>, we can suppose that aspects like

- availability of a technician in case of technical problems,
- the interaction with team mates working in the same interpreting booth or across booths in multilingual conferences,
- the (sometimes very limited or even non-existing) possibility of talking to speakers and audience before the conference or during breaks,

---

<sup>13</sup>Note that, apart from the interpreter’s console with standard control buttons for the incoming sound, an interpreter in a presence-based conference setting also has only very little control of the actions in the room.



- self-control of the direction of sight, focusing on speaker, audience, presentation or any other visual input available

are factors that interpreters feel they need to control.

The design of an interpreting hub as described above could easily cover the first two aspects, as such a hub would necessarily need the presence of a technician onsite in the hub to manage the technical equipment and the whole team (if interpreting takes place from one hub only) or at least the booth mate would be working together at the same place. The interaction and communication with speakers and audience would, of course, require a specific solution, but could be handled e.g. considering ‘institutionalized’ briefing and Q&A sessions before, during and after the meeting.

Control of the view on the speaker, audience and additional visual input such as presentations, in presence-based interpreting usually performed by head movements and eye focusing, would need further development based on the technologies described in §6.1. Considering the very fast development of image captioning, transmission and reproduction, it seems only a matter of time until these technologies will be affordable and experience an adaptation to the needs of remote simultaneous conference interpreting.

## 7 Conclusions

In times of growing demand for flexible, accessible and customer oriented digitalized communication services in a globalized world, technological solutions for quality simultaneous interpreting services will have to be developed, taking into account several aspects related to the organization and delivery of those services. Existing standard solutions developed for other purposes, such as video conferencing or web conferencing without interpreting, or solutions for consecutive interpreting in specializations other than conference interpreting, appear not to be sophisticated enough to meet the special requirements that distance interpreting imposes in terms of sound and image input to interpreters in bilingual and multilingual conference settings. Technological enhancements in the field of virtual reality and augmented reality, as well as immersive communication environments may offer the possibility to overcome existing constraints.

One of the major challenges for interpreting studies in the field of distance interpreting will be to find a more interdisciplinary and future oriented approach, building teams of researchers in the technical, medical and psychological field, to name only a few of them, and to combine these different research disciplines

in multidisciplinary projects that can actively lead to designing the future workspace for conference interpreters, in the first place, but for other specialized interpreting services as well. It goes without saying that fellow (conference) interpreters need to be prepared for remote simultaneous interpreting during their training, as this modality is experiencing a growing demand in different interpreting specializations, including tele- and videoconference interpreting (Braun2015)■ Apart of the integration of training modules designed specifically for this modality, addressing cognitive, communicative and technical aspects, amongst others, this would also require adequate equipment of training facilities with the appropriate features (audio/video conferencing hard- and software and connectivity, to mention just a few of them).



# Did you like this book?

This book was brought to you for free

Please help us in providing free access to linguistic research worldwide. Visit <http://www.langsci-press.org/donate> to provide financial support or register as a community proofreader or typesetter at <http://www.langsci-press.org/register>.





# Interpreting and technology

Since the 1970s, the notion of a lexeme, an abstract lexical unit identifying what is common to a set of words belonging to the same inflectional paradigm, has become a cornerstone of theoretical thinking on morphology and a standard tool for description. The present volume collects papers that crucially use, discuss or question the lexeme in the context of contemporary morphology, with particular emphasis on its place in the description of word formation through the concept of a *Lexeme Formation Rule*. It will be of interest to any descriptive linguist, theoretical linguist, or psycholinguist with an interest in morphology and its interface with syntax and lexical semantics.

