The Interpreters' Newsletter 2023 (28), 69-90
DOI: 10.13137/2421-714X/35551 ISSN 1591-4127 (print) ISSN 2421-714X (online)
https://www.openstarts.units.it/dspace/handle/10077/2119

Enhancing numerical accuracy in simultaneous interpreting students: an exploratory study comparing human and AI-based support

Claudio Russello
Università degli Studi Internazionali di Roma, UNINT

Matilde Carbutto
University of Surrey

Abstract

Simultaneous interpreting (SI) of numerical elements is a complex task and numbers are widely acknowledged as a problem trigger and a source of errors for interpreters. However, the latest advancements in Natural Language Processing (NLP), Artificial Intelligence (AI) and Automatic Speech Recognition (ASR) integrated into Computer-Assisted Interpreting (CAI) tools have shown potential in enhancing accuracy during SI of numerical elements, as well as of other parts of speech, such as specialised terminology. Although many recent studies are focusing on this area, research comparing the impact of CAI tool support during SI against the support provided by human boothmates is scarce. This paper aims to address this gap within this emerging field of research by comparing the performances of interpreting students using the ASR-powered CAI tool InterpretBank with those who worked with the support of a human boothmate while interpreting a source text (ST) with a high density of numbers. The pilot study assesses the accuracy and intelligibility rates of numerical renditions obtained by novice interpreters. The findings show that CAI tools can improve interpretation quality, and the digital boothmate provides equivalent support to a human boothmate in terms of accuracy of numerical rendition, while considering latency limitations.

Keywords

Computer-Assisted Interpreting (CAI), Automatic Speech Recognition (ASR), Artificial Intelligence (AI), interpreting technologies, digital boothmate, *InterpretBank*, numbers

The field of conference interpreting is a broad and diverse area of study that is constantly evolving. The rapid technological advances affecting many language professions are having a significant impact on the field of conference interpreting as well, shaping new training paradigms, transforming the job market, and changing the relationship between interpreters and the technology they use on a daily basis. These premises make it easy to comprehend the increasing interest of the scientific and professional communities in the various manifestations of possible human-machine (or interpreter-machine) interactions.

Undoubtedly, technology has always played a fundamental role in the field of interpreting, e.g. by enabling the emergence of SI after the Second World War, and over the years it has significantly revolutionised the methods used to carry out interpreting assignments.

In addition to the technologies enabling Remote Interpreting (RI), which have been transforming the settings in which interpreters work since the 1970s (Baigorri-Jalón 2014), the last two decades have witnessed the emergence and proliferation of CAI tools (Fantinuoli 2018a; Braun 2020). While the first generation of CAI tools was primarily designed to support interpreters in the preparation phase (before the event), subsequent generations have led to the more sophisticated CAI tools available today, which offer a wealth of possibilities that can assist professionals at every stage of their assignments (including during the event).

In fact, the advances in the field of NLP, AI and ASR have paved the way for the development of new CAI tools that can directly assist interpreters in their work by proposing in real time the automatic transcription and translation of the full ST or of the main problem triggers extracted from the speaker's speech, such as numbers, named entities and complex terminology, through suggestions that interpreters can consult directly on their monitor.

Among these tools, *InterpretBank* (Fantinuoli 2017), the software used for this study, includes a feature called Artificial Boothmate, which integrates ASR systems into a user-friendly web-based interface for interpreters. It should be noted that there are other technological solutions emerging in the interpreting industry that offer similar real-time human-machine interaction support such as *Cymo Note* (Goldsmith 2023), *KUDO Interpreter Assist* (Fantinuoli *et al*. 2022), *SmarTerp* (Rodriguez *et al*. 2021), to name but a few (see §1).

It seems reasonable to assume that the proliferation of these technological solutions and the growing interest of professionals in the possibility of receiving suggestions from a digital boothmate, in contrast to their initial reluctance (Fantinuoli 2018b), are motivated not only by technological advances that now make the available tools reliable enough to be used in interpreting, but also by the significant increase in RI. As evidenced by recent studies (Raaya/Martin 2022) Remote Simultaneous Interpreting (RSI) is no longer an incidental or sporadic activity, as it may have been before the outbreak of the Covid-19 pandemic or during the subsequent lockdown period but is now fully integrated into the normal workflow of many interpreters.

Indeed, although many interpreting delivery platforms (IDPs) now have built-in features specifically designed for interpreters (e.g. for managing handovers or pri-

vate communication through a dedicated chat), having a booth partner who is remote means that the teamwork is very difficult to maintain (Ordeig 2019). As working remotely becomes more prevalent, the absence of a physical booth partner poses a challenge for interpreters. Therefore, while the possibility of relying on a boothmate is diminishing, the availability of automated solutions that can provide interpreters with the suggestions traditionally offered by a boothmate is increasing.

This paper presents a pilot experimental study designed to explore the effectiveness of CAI tools with integrated ASR and AI in aiding novice interpreters with accurate and intelligible interpretation of numerical data in STs.

The following two sections provide an overview of the state-of-the-art CAI tools available today and the related studies already conducted in this area. Then, the experimental setup and materials are presented, together with the accuracy results obtained by the control group, which interpreted a text with a high density of numbers with the help of a boothmate, and by the experimental group, who worked with the support of the software. The study will also analyse the responses to an evaluation questionnaire provided to a sample of independent listeners who assessed the overall intelligibility and impact of pauses in the renditions.

1. Computer-assisted interpreting tools

One of the first publications in the field of Interpreting Studies that explored the possible use of software specifically designed to assist interpreting is that of Rütten (2004). The author presented a comprehensive theoretical model of CAI tools, envisioning a software prototype that could assist interpreters at each stage of the process. This includes a language-oriented terminology module, a content-oriented documentation module, and a general situation-oriented module. This concept served as the basis for the development of CAI tools, which were initially limited in scope and focused on managing bilingual glossaries. Since then, they have evolved to become a resource for assisting interpreters in a more holistic manner.

InterpretBank is one of such comprehensive CAI tools that was developed between 2008 and 2012 at the University of Mainz/Germersheim. The tool is designed to serve as an efficient workstation that facilitates interpreters in streamlining their workflow before and during an event. The software offers a range of features that support interpreters, including machine translation of the glossary's entries, terminology collection from online resources, seamless integration of preparatory material, automatic extraction of relevant specialised terminology, concordance searches, and a memorisation interface for learning the terminology contained in the glossaries. In terms of the functionalities that can be employed during an assignment, the software encompasses two key features: a manual query function that enables users to conduct manual searches within the glossary, and an ASR-based functionality. The latter, namely the Artificial Boothmate feature, is a cloud-based function that automatically searches for terminology, proper names, and numbers during interpretation by listening to the original speaker, displaying in real-time on the interpreter's screen all the numbers pronounced by the speaker and the terms present in the glossary in use.

InterpretBank has been the subject of several studies testing its performance in SI, and most of them seem to agree that the software helps interpreters render terms

and numbers more accurately (Prandi 2015, 2018; Guo *et al.* 2023: 94; Pisani/Fantinuoli 2021; Fantinuoli *et al.* 2022). However, there is some disagreement among SI scholars. For instance, Zhou (2019 in Guo *et al.* 2023: 94) found that although *InterpretBank* may improve overall performance, some participants' renditions were negatively affected.

Advancements in technology have resulted in the creation of further advanced CAI tools that aim to cater to all the needs of interpreters, offering a range of features such as remote interpreting, terminology management, and live support during assignments. Among the many advanced CAI tools of this kind available, this section will provide an overview of *Cymo Note, Kudo Interpreter Assist and SmartTerp*, as referenced above, as they can assist interpreters specifically with numbers pronounced by speakers, which is pertinent to our research.

Cymo Note is a professional multilingual note-taking software designed to provide automatic speech recognition in various interpreting settings, including remote, onsite, and hybrid meetings, and modalities such as simultaneous, consecutive, and hybrid approaches. The tool allows users to add names and terms to the glossary and highlights them in the running transcription. *Cymo Note* offers a pay-per-minute pricing system based on the type of speech recognition engine used, including *Microsoft ASR*, *Tencent Speech-to-text*, *iFlytek Speech-to-text*, and the proprietary *Cymo Speech Engine*. The software supports over 15 languages and allows users to build their glossary and copy and paste entire glossaries for repeat clients. Furthermore, the consecutive interpreting feature of *Cymo Note* is its most innovative aspect, with its ability to divide the screen into two, transcription on the left and a blank space for taking notes on the right, using the drawing mode to take notes anywhere on the page (Goldsmith 2023).

Kudo Interpreter Assist has emerged as a popular CAI tool in recent years, designed specifically for integration into RSI scenarios. Its primary objective is to reduce preparation time and improve the accuracy of interpretation in highly specialised events (Fantinuoli *et al.* 2022). *Interpreter Assist*'s salient features comprise an automatic glossary creation tool and a real-time suggestion system. Utilising this tool from a user's standpoint involves creating a project, identifying the event topic, optionally assessing the automatically generated multilingual resources, sharing resources with team members, and conducting RSI with real-time suggestions. From the machine's perspective, the process encompasses the creation of a mono or multilingual domain-specific corpus, extraction of monolingual terminology from the corpus, translation of the terminology into target language(s), fine-tuning of baseline models with generated resources, optional fine-tuning with client-generated information, and automatic access to resources in real-time via the interpreter console.

Finally, *SmarTerp* is an EU-funded Innovation Activity aimed at developing a remote simultaneous interpreting platform with an ASR- and AI-powered CAI tool. The RSI system aims to provide high-quality RSI services with ISO-compliant audio and video quality, communication options with all actors involved in the assignment, and an RSI console that allows interpreters to perform all key actions required by this interpreting mode (Frittella/Rodríguez 2022). The ASR/CAI tool employs sophisticated algorithms to transcribe spoken language into written text, supporting interpreters in rendering common problem triggers such as named entities, acronyms, specialised terms,

and numbers. The CAI tool has undergone significant development, and ongoing research and development activities related to system performance and accessibility are improving both *SmarTerp* components. Overall, *SmarTerp* is a promising CAI tool that has the potential to enhance the quality of interpreting in various settings (*Ibid.*).

In general, CAI tools are already revolutionising the interpreting field by facilitating easier and more efficient work for interpreters, as they offer valuable support and can enhance the quality of interpreting in various contexts. The objective of the present pilot study is to focus on the assistance offered by CAI tools for numbers to interpreting trainees. The rationale for focusing on the analysis of number management during interpretation stems from the widely shared belief among professional interpreters and academic experts in the field regarding the criticality of handling numerical elements during SI (see §2). This criticality is reflected in a high error rate, as will be discussed in the subsequent section. Therefore, other applications of interpreting technologies, including remote interpreting, corpus-driven preparation, computer-assisted consecutive interpreting (CACI), computer-assisted interpreters' training (CAIT), and hybrid interpreting modalities, among others, will not be considered in this study.

2. Related studies

There appears to be a general consensus among professional interpreters and scholars in the field regarding the criticality of dealing with numbers in a ST as actual problem triggers during SI (Gile 2009; Setton/Dawrant 2016; Frittella 2019). As a result, they are often the cause of declining quality in SI, with error rates ranging from 21% to 70%, as shown by several empirical studies focused on this topic (Braun/Clarici 1996; Lamberger-Felber 2001; Mazza 2001; Pinochi 2009; Desmet *et al.* 2018; Collard 2019; Frittella 2019; Kajzer-Wietrzny *et al.* 2021). Despite the critique offered by Frittella (2022: 33) regarding the calculation of error rates and the ambiguity surrounding the definition of "number," as well as the potential impact of methodological variations in some of those publications, it is undeniable that there is a growing interest in computer-assisted SI research. This interest is driven by both scientific and academic, as well as technological and market considerations, with a focus on techniques and specific approaches to address these disruptions through interpreter-machine interaction (Desmet *et al.* 2018; Canali 2019; Defrancq/Fantinuoli 2020; Pisani/Fantinuoli 2021, Guo *et al*. 2023).

Within the experimental studies on interpreters' performance relating to number rendition in SI, one factor that seems to play a crucial role is the presence, or absence thereof, of some form of visual support during the interpretation, regardless of whether it consists of: A) the note jotted down by the interpreter or the boothmate as soon as he or she hears a number, B) a text provided to the interpreters with the speech to be interpreted or C) a suggestion shown shortly after the speaker pronounces the number provided by a technological support based on ASR.

Based on the scientific literature, it seems clear that the first two cases have noticeable advantages. Lamberger-Felber (2001) reported an increase in accuracy of approximately 50% when interpreters are provided with a text, and Mazza (2001: 96) observed an accuracy increase of up to 10% when a human boothmate is present.

Additionally, two corpus studies based on the EPTIC corpus, which includes speeches from the European Parliament and their interpretations, found varying error rates for numbers. In one case, the error rate was around 18% (Collard/Defrancq 2017), and in another, it ranged from 4 to 7%, albeit with a 20% omission rate (Kajzer-Wietrzny *et al.* 2021). These error rates are much lower than the average of previous studies and far from the peak value of 70% reported by Braun and Clarici (1996). Regarding these results, however, in Desmet *et al.* (2018: 15 it is pointed out that the difference between experimental and corpus results may be attributed to the presence of a colleague in the booth assisting the interpreter with taking down numbers. This assistance is often not available in experimental settings, as participants usually work alone in the booth.

However, while assistance and visual input have been found to enhance performance, these factors are often beyond the control of the interpreter. To overcome this limitation, interpreters have today the opportunity to get help from the software that displays the relevant numbers on a screen in the booth immediately after they are mentioned. In the following, we will present a diachronic analysis of experimental research on this approach. The results obtained so far have been quite promising.

The increasing research interest in the analysis of interpreter-machine interaction through automatic software querying research tools started with the introduction of the first prototype CAI tool with integrated ASR software for interpreters by Fantinuoli (2017). The results of the empirical study in question aimed - at this pioneering stage - at evaluating the performance of the experimental Artificial Boothmate functionality of the *InterpretBank* software, without yet investigating the possible results of the interpreter's use of the CAI tool. The study showed that, although ASR engines were not yet perfectly adequate and presented problems in certain circumstances, such as in listening to non-native accents, managing sectoral terminology, and others, they could already be successfully used as a technology for SI. Furthermore, Fantinuoli (2017) emphasised the need for further studies to determine whether the support of ASR-powered CAI tools could increase the overall quality of interpreters' performances.

Subsequently, further experiments on the rendition of numbers using mock-up systems, which involved the projection of the decontextualised numbers contained in the ST on the interpreters' computer screen immediately after the numbers were pronounced by the speakers, confirmed a noticeable reduction of error rates from 43.5% to 13.5% (gain 30%) in Desmet *et al*. (2018) and from 64% to 25% (gain 39%) in Canali (2019). It must be emphasised here that experimental designs with mock-up systems are not based on the actual automatic recognition of numbers in a given speech. ASR systems do not achieve (yet) perfect recognition and minimal software latency (the delay between a user's action and a web application's response to that action) as the mock-up system does (see §3.5). Therefore, the results describe the maximum performance that could be achieved with an ideal system, in terms of accuracy and even more so latency.

A turning point was reached with a new experiment conducted by Defrancq and Fantinuoli (2020), in which they applied a real output produced by the CAI software in real-time and a different visualisation approach: the numbers were incorporated into the complete transcription and highlighted, recording a reduction in the error rate from 32.3% to 9.8%.

More recently, there have been several lines of research dealing with the rendering of numbers in computer-aided SI, with a particular focus on improving CAI tool suggestions. Brunetti (2023), for example, conducted an experimental investigation on the "spillover effect", a frequently observed issue when rendering number-accompanying items, by designing an experimental mock-up support interface. The study involved SI from English to Italian and vice versa of the same speeches using two different supports. The control group used *InterpretBank*'s Artificial Boothmate feature, displaying only the figures mentioned and some lexical items, mainly adverbs of time. On the other hand, the experimental group used the mock-up support, which provided both figures and their referent, as well as the use of punctuation to delimit the order of magnitude of the numbers in the ST. The experimental support led to an accuracy increase of 28.5% EN>IT interpreting and 28% in IT>EN interpreting, confirming the initial hypothesis that providing the lexical elements after the number up to the punctuation mark could be an effective solution to the "spillover effect". In this study, several categories of errors were considered to assess the participants' performances, referring to criteria such as syntactic, lexical, approximation and omission errors. Additionally, the hypothesis that providing numbers including punctuation increases the comprehensibility of the numbers and, consequently, the accuracy of their interpretation was also confirmed. It is important to note that this study did not feature software latency or misrecognition, as specified for the mock-up system used by Desmet *et al.* (2018). The study also proposed a modification regarding the selection of numerical elements to display as suggestions, eliminating simple numbers (one, two) or commonly repeated contents, such as in the case of Covid-19.

As a matter of fact, one of the main limitations of such systems is to be found in the lack of selectivity in the choice of suggestions proposed on the interpreter's screen. Especially in the case of numbers, the subject of the present research, no distinction is made between complex and simple ones, which leads the system to display numbers such as 100 and 153,867 indiscriminately, even though these have different levels of difficulty (Pisani/Fantinuoli 2021: 184). This entails several disadvantages, such as the overabundance of unfiltered suggestions received by the interpreter, who might therefore experience a cognitive overload, a possible cause of an overall decrease in the quality of the rendition.

One additional limitation that emerges from the aforementioned empirical studies is the impact of software latency on hint display. The delay caused by system latency affects the ear-voice span (EVS), as it influences the time interpreters have to wait before incorporating suggestions into their delivery. This delay can often render CAI support ineffective and even undermine the performance of the interpreters. For instance, Zhang (2020 in Guo *et al.* 2023: 94) found that consecutive interpreters who use ASR experience significantly higher latency and are generally less fluent and more prone to false starts, hesitation, and filler words. An empirical study conducted by Fantinuoli/Montecchio (2023) found that interpreters can incorporate suggestions by temporarily expanding their EVS for up to 2 seconds without any negative impact on their performance. They can even extend it to 3 seconds without significant disruption. However, if the system latency exceeds this point, the precision of such suggestions is severely compromised, and there is a risk of losing important information in the interpretation. These findings align with the study by He (2018 in Guo *et al.* 2023:

94), which found that delaying ASR results by more than 5 seconds negatively affects performance.

In addition to numerical support, it is worth noting that several experimental studies have been conducted on automated lookup tools for specialised terminology (Prandi 2015,2018,2022 as well as Biagini 2015; Zhou 2019 and Zhang 2020 in Guo *et al.* 2023: 94). These studies have indicated the potential to enhance the final output quality, albeit with modest improvement due to additional cognitive load and perceived distraction during user-machine interaction. However, as this research area is beyond the scope of the current study, it will not be further explored.

3. Exploratory study on the use of CAI/ASR vs human support in SI

In order to explore the current state of CAI tools powered by ASR, specifically the effectiveness of the experimental Artificial Boothmate feature of *InterpretBank* when compared to a human boothmate, a study was conducted with a group of non-professional interpreters[1]. Its main objective is to determine to what extent such features can be integrated into the workflow of simultaneous interpreting trainees in a functional manner, particularly when dealing with numerical information. Additionally, the study aims to assess the potential of the systems under investigation in conceivably substituting traditional interpreting assistance techniques, such as the presence of a boothmate.

The following sections will provide details on the methodology used in the study, by introducing the participants (§3.1), the experimental materials and setup (§3.2 and §3.3), the evaluation criteria applied to assess performance (§3.4), followed by the results of the ASR system (§3.5) and the methodological limitations (§3.6). Finally, Section 3.7 and 3.8 will analyse respectively the accuracy and the intelligibility results obtained.

3.1. Participants

The experiment involved 16 second-year students who were enrolled in the Master's Degree course in Interpreting and Translation at Università degli Studi Internazionali di Roma – UNINT, with 8 participants each in the control and experimental groups. Additionally, 8 more participants took part in the experiment as confederates acting as boothmates (see also Chmiel/Spinolo 2022) for students in the control group. All participants had completed the EN-IT introductory course in simultaneous and consecutive interpreting during their first year, followed by the EN-IT simultaneous and consecutive interpreting course during their second year. The reason for including students instead of professional interpreters was because they were a homogeneous group with comparable overall training in interpreting and confirmed knowledge of the software utilised, having completed a course that covered new technologies for

1 This article is grounded in Carbutto's Master's thesis (2022), functioning as both its fundamental basis and primary source of collected data.

conference interpreting, with a particular emphasis on CAI tools and their integration into SI. However, it is possible that most participants had not previously applied their skills in this area in a professional setting. Moreover, no advance preparation for computer-assisted interpreting was necessary, and no information about the software to be used was provided before the experiment.

For this exploratory study, a between-subject design was chosen to maintain consistency and minimise variations resulting from the use of different STs, although within-subject designs are typically preferred to account for individual differences, especially with a small sample size. This allowed to test an experimental procedure that could then be applied to larger groups of participants.

3.2. Experimental materials

In order to analyse the rendition of numbers in the context of assisted SI, a report on labour market indicators during the Covid-19 pandemic from the European Commission's Directorate-General for Employment, Social Affairs and Inclusion was selected. The original text[2] was edited to include complex numbers (cf. Appendix A). A native English-speaking interpreter provided a voice recording of the resulting text, with an average delivery rate of approximately 118 words per minute, considered adequate for SI purposes by previous studies (Seeber 2015: 85), as it does not noticeably impair the interpreter's performance. The speaker prosodically emphasised specific terms and passages deemed relevant, with numbers pronounced clearly and slightly slower than the rest of the speech. The final text contains 1,082 words, with 64 of those being numerical, and the audio recording has a total duration of 9:19 minutes.

3.3. Experimental setup

Prior to starting the experiment, the participants were briefed on the procedure and five minutes before the experiment, they were provided with a glossary in paper format containing 17 technical terms that they could consult during the interpretation. Each control group participant was randomly paired with one of the eight confederates acting as boothmates who could assist the student in transcribing numbers. The boothmates were instructed to only note down the numbers without suggesting terminology translations or replacing the interpreter.

To ensure uniform conditions for all experimental group participants, a single-screen recording was used instead of allowing each interpreter to use their own PC. This recording featured the audio of the speech to be interpreted and *InterpretBank*'s Artificial Boothmate. The two windows had been arranged on the screen next to each other in such a manner that the terminology column in the software window was not shown, as it was not relevant to the experiment under study, as shown in Figure 1.

2 Joint Employment Report 2021, as adopted by the Council on 9 March 2021: <https://ec.europa.eu/social/BlobServlet?docId=23156&langId=en> (last accessed 12/12/2023).

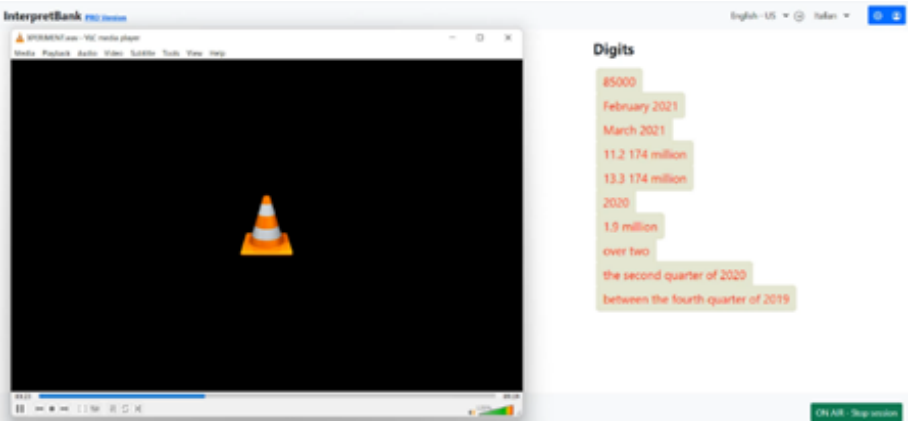

Figure 1: Sample of the screen provided to the experimental group

After a 5-minute warm-up session to help the interpreters get settled and successfully test the system, they were asked to record their interpretation using a mobile device and were assured that the recordings would be used anonymously for the study. To this end, participants were administered written informed consent, outlining the study purpose, data collection procedures, and steps taken to protect anonymity. Detailed information on data treatment, including the removal of personally identifiable information, was also provided. Following the interpretation, the interpreters were asked to submit their recordings and complete a questionnaire about their performance and booth support.

3.4. Evaluation criteria

As early as 1982, AIIC described interpreting quality as "that elusive something" that "no one can define", making it one of the most extensively researched topics in conference interpreting. Numerous researchers, including Chevalier and Gile (2015), have postulated that both "informational fidelity" and "presentation" are quality components. Therefore, a "hybrid" quality assessment method was chosen for this study to measure both intelligibility and fidelity. Intelligibility was assessed using the intelligibility scale originally proposed by Carroll (1966) and later adapted for interpreting by Tiselius (2009), using the same approach adopted by Pisani and Fantinuoli (2021) to compare speech-to-text translation and human interpreting. In addition, two extra questions were included in the questionnaires provided to the evaluators so that they could provide their assessment of the impact of pauses in the renditions of the participants being tested.

For content analysis and the assessment of the degree of informativeness of the target texts (TT) with respect to the ST, a method entirely focused on the rendering of numbers in SI was adopted. This method has been used in similar studies, such as those of Braun and Clarici (1996), Pinochi (2009), Desmet *et al.* (2018) and Pisani and Fantinuoli (2021). For the evaluation of the TTs and the classification of errors, we applied

the error categories defined in similar studies on the interpretation of numbers (Braun/Clarici 1996; Pinochi 2009; Pisani/Fantinuoli 2021). The error typologies considered were the following: (1) omissions, where the number is either omitted or replaced by a generic expression like "some"; (2) approximations, where the number is rounded off; (3) lexical errors, where the order of magnitude is correct but one or more digits are misinterpreted, for instance, 35.3 instead of 35.5; (4) syntactic errors, where the number is of the wrong order of magnitude, for example, 15 thousand instead of 15 million; (5) phonological errors, where the error is believed to be due to a phonetically incorrect perception of the number, like 65 million instead of 16.5 million; (6) inversion errors, where the digits are correct but pronounced in the wrong order, such as 19 million instead of 91 million; and (7) other types of errors, which is a macro-category that includes all errors that do not fit entirely into any of the other categories.

For the purpose of this study, 16 tables were filled out, one for each interpretation, as shown in Table 1. In the first two columns on the left, the numbers pronounced by the original speaker and the numerical class of each number were transcribed. Then, the numbers as pronounced by the interpreters in the TT were included. The last three columns indicated whether the number was interpreted correctly, and, if possible, categorised the type or types of error made by the interpreter. For the analysis of the renditions of the participants in the experimental group, two additional columns indicated whether the number was successfully transcribed by the software or not, and the actual transcription.

Number in ST	Numerical Class	Number in TT	Correct	Type of Error	Type of Error	Transcribed	Transcription
2021	K	2021	TRUE			TRUE	2021
2013	K	/	FALSE	Omission		FALSE	/
6.2 million	M	6,2 milioni	TRUE			FALSE	/
4th	U	1°	FALSE	Lexical		TRUE	Fourth
2019	K	/	FALSE	Omission		TRUE	2019
2nd	U	2°	TRUE			TRUE	Second
2020	K	2020	TRUE			TRUE	2020
2	U	/	FALSE	Omission		TRUE	2
1.9 million	M	/	FALSE	Omission		TRUE	1.9 million
2020	K	2020	TRUE			TRUE	2020
13.374 million	M	13%	FALSE	Syntactic	Approxi-mation	TRUE	13.374 million
11.274 million	M	174 milioni	FALSE	Phono-logical		TRUE	11.2 174 million
2021	K	2021	TRUE			TRUE	2021

Table 1: Table for evaluation

In accordance with the approach adopted by Pisani and Fantinuoli (2021), instances where multiple types of errors were identified for the same number were treated as a single error, despite the distinct types being indicated in the respective error type columns. Consequently, discrepancies arose between the total number of errors and the number of errors categorised by type.

3.5. ASR results

The ASR system transcribed 56 of the 64 numbers contained in the ST, which accounts for 87.5%. Of those 56, 51 were transcribed completely accurately, representing 79.7% of the total numbers in the text.

Upon scrutinising the software latency, it was observed that the mean latency stood at 4.2 seconds, (*SD*= 3.9). The data exhibited a median latency of 2.67 seconds. The recorded latencies ranged between a minimum of 1 second and a maximum of 18.75 seconds, highlighting the variance within the system's performance. The software exhibited increased latency or even failed to display the number on screen in the more discursive sections of the ST, which had a lower density of numbers. However, in the sections with a high concentration of numbers, the software was able to transcribe and return them on the screen correctly, with minimal latency.

Despite the rather high accuracy, the primary disruptive element and source of difficulty associated with the software lies precisely in the latency factor. In addition to the omissions made by the ASR system, the most significant errors were found in the high numerical class numbers, such as those on the order of millions and complex numbers, like "1.931 million", transcribed as "1.9 130 1 million" in its final form, possibly because some numbers were pronounced somewhat slower than the rest of the text (§3.2), causing considerable difficulties for the interpreters. Finally, it is worth noting that in correspondence with time references, the software transcribed not only the year, such as "2019", but also a portion of the context of use, such as "in the last months of" and this is likely to have increased the interpreters' cognitive load as they were faced with more information to process visually and seamlessly integrate into their oral workflow.

3.6. Limitations

As it is always the case, our exploratory study is not without its limitations that need addressing. The primary constraints include the restricted participant pool and an experimental design that omitted the collection of handwritten notes from human boothmates. The inclusion of such notes in future studies could provide valuable insights for comparing the numbers transcribed by the CAI tool with those recorded by human boothmates, facilitating cross-comparisons with the interpretations given by the experimental and control groups in those specific circumstances.

Moreover, it should be borne in mind that the participants in our study were novices, and it's conceivable that different results may manifest in a similar study involving expert interpreters.

These acknowledged limitations highlight the necessity for additional exploration in more specialised settings to validate our initial findings and to advance the research in the field of human-machine interaction within the broader context of interpretation and technology.

3.7. Accuracy results

This contribution aims to evaluate the effectiveness and practical application of ASR-enhanced CAI tools to potentially replace the need for human support in interpreting numerical information during SI.

Upon analysing the data, it was found that the control group, who worked with a boothmate, had an error rate of 55.6%, while the experimental group, who utilised the ASR software, had an error rate of 50%. It is worth noting that the underlying cause for such a low degree of precision can be attributed to the significant concentration of numbers in the ST, which had been revised to be suitable for experimental purposes, coupled with the fact that participants were novices.

After a more fine-grained analysis of the error types made by each group, it became clear that omissions were the most common error committed by both groups. Other types of errors were also present, but they were distributed fairly evenly. A comprehensive overview will be presented below, accounting for the various numerical categories and the corresponding frequency of each error type.

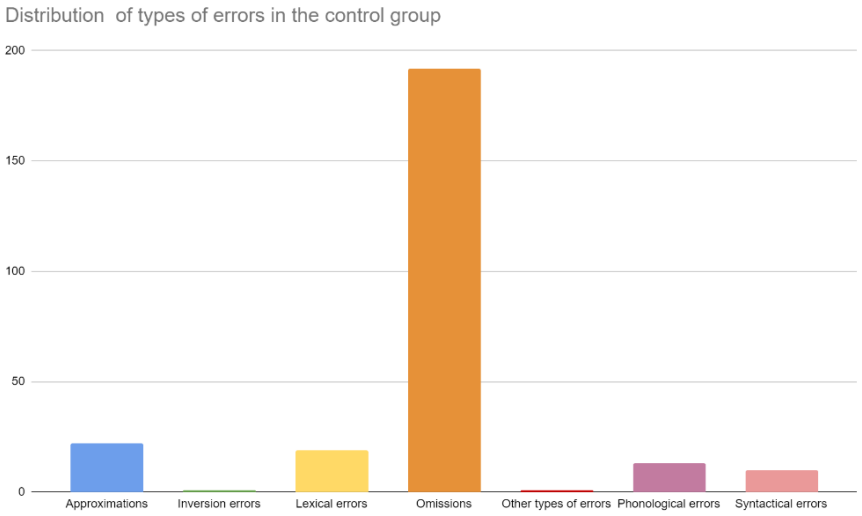

Figure 2: Distribution of types of errors in the control group

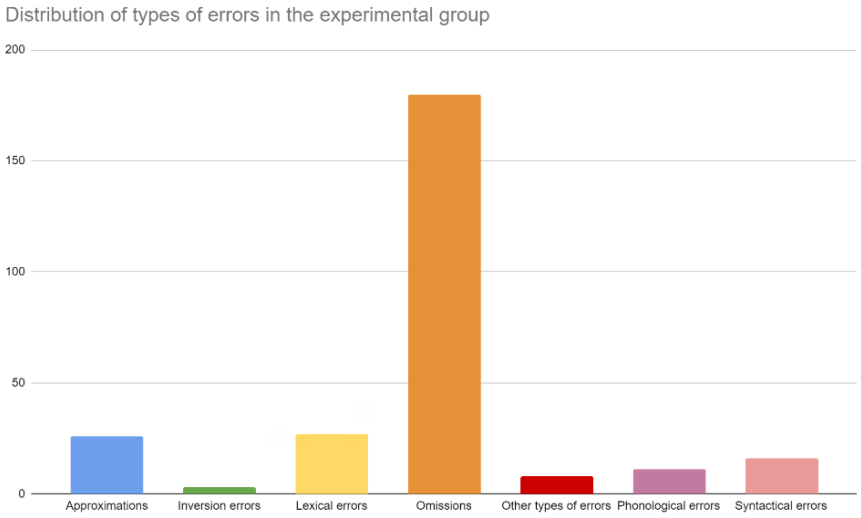

Figure 3: Distribution of types of errors in the experimental group

As can be seen from the Figures 2 and 3, the most common error is omission, in fact for the control group out of a total of 249 errors, there were 192 omissions, while for the experimental group out of a total of 257 errors, 180 were omissions. It is important to note that errors categorised as omissions, whether total or partial, can be construed as either outright errors or as strategies. Frequently, interpreters opt to strategically omit the pronunciation of numbers to avert cognitive overload and, thus, to avoid compromising overall performance, despite the concomitant sacrifice of numerical precision. Indeed, as retrieved from the analysis of the recorded performances, this phenomenon occurred in the control group particularly in those parts of speech that contained a higher density of numbers, while in the experimental group, the phenomenon was attributable to the failure of the computer system to transcribe the numbers or to latency times that were too prolonged for interpreters to avail themselves of the tool's support by pronouncing the numbers in the TT. This outcome for the experimental group also seems to corroborate the notion of participants exhibiting an over-reliance on the tool, a phenomenon that has been documented in previous research (Prandi 2015; Defrancq/Fantinuoli 2020; Frittella 2022). Although, regarding the distribution of omissions per numerical class, there seem to be no remarkable differences between the two groups, it is noteworthy that omissions constituted 75% of the total errors in the control group and 66% in the experimental group, underscoring the software's effectiveness in mitigating a type of error that could substantially diminish the overall informativeness of a TT.

Regarding the other types of errors, approximation emerges as a prominent strategy utilised to a considerable extent by interpreters in both the experimental and the control group. It constitutes 8.9% of errors in the control group and 9.6% in the experimental group. Although the control group appeared to have a better overall performance, such results may actually point at a possibly higher effectiveness of computer

assistance in the interpreting booth. As a matter of fact, the participants of the control group working with a boothmate often avoided translating particularly complex numbers like "11.274 million" or "1.931 million", possibly due to their boothmate's inability to prompt them. In contrast, those who employed the software succeeded in at least partially expressing the number, resorting to solutions such as "11 million" and "almost two million", despite the system's erroneous transcription ("11,2 174 million" and "1.9 130 1 million"). Furthermore, the single-digit numbers (e.g. "over the past 5 years") approximated by participants in the control group, utilising phrases such as "some" and "in recent years" were mostly reported correctly in the experimental group.

Examination of the dataset produced an unforeseen insight with regard to the prevalence of lexical errors (19 in the control group and 27 in the experimental group) and syntactic errors (10 in the control group and 16 in the experimental group), which together constituted 11.4% of errors in the control group and 16% in the experimental group. Syntactic errors, which have more serious implications as they result in the translation of a much larger or smaller number, occurred only 16 times among all participants in the experimental group, averaging to 2 syntactic errors per interpreter. This represents only a small fraction of the 64 numbers included in the original speech. Upon deeper analysis, it became evident that, in the experimental group, this type of error almost always occurred in conjunction with incorrect or incomplete transcriptions by the software, such as the above-mentioned instance of "1.931 million". Additionally, in half of the instances, the aforementioned error occurred when dealing with numerical values in the millions range, whereas in 7 cases out of the total number of 16, the error occurred while processing numbers that included a decimal point.

Conversely, inversion errors were not a significant finding in either group. The only instances in which inversion errors occurred, a total of two occurrences, were observed in the renditions of the experimental group, specifically when dealing with complex and million-order numbers, as in the case of 13.374 million interpreted (in Italian) as 31.174 million. This was an anticipated outcome, given that the language of the source speech, namely English, has a linear numerical system that corresponds to the Italian system. The likely reason for these errors, as revealed by the analysis of the screen recording, is the system's delay in transmitting the transcripts to the interpreters' screens.

Likewise, the "other typology" errors, which fall outside the established error categories and are limited to the experimental group, are not significant and therefore do not require in-depth analysis. In fact, the inaccurate transcriptions produced by the software were the main cause of the varied errors committed by the interpreters. These findings emphasise the pressing requirement for enhanced accuracy in the software's transcription of complex and large numbers (e.g. 11.274 million, 2.359 million and 1.931 million in our ST). It is these particular types of numbers that present the most significant challenges and cognitive load for simultaneous interpreters (Frittella 2019; Mazza 2001). Consequently, it is crucial for the software to offer reliable and precise support specifically for complex numbers, given that they are the most demanding to comprehend, process and interpret.

On the contrary, the phonological error distribution, which accounted for 5% of the total errors in the control group and 4% in the experimental group, provides ev-

idence of the software's validity in transcribing simple and decimal numbers, even when included in a dense list of numerical information. As a matter of fact, the participants who used the ASR system made fewer errors in phonetic perception, with an average of 1.6 errors per participant in the control group and 1 error per participant in the experimental group, presumably due to the more effective transcription of numbers by the ASR software compared to the transcription by boothmates in the control group.

3.8. Intelligibility results

Drawing on the evaluation theories proposed by Tiselius (2009) and Carroll (1966), 12 native Italian speakers between the ages of 24 and 69 with no prior training or experience in the field of SI were involved as evaluators. Each evaluator was asked to listen to four short extracts of four different performers, two belonging to the control group and two to the experimental group, totalling approximately three minutes in total for each performer. Each extract was therefore listened to and evaluated by 3 different evaluators, according to the following scheme:

EVALUATOR 1	PART A1	PART A8	PART B1	PART B8
EVALUATOR 2	PART A2	PART A7	PART B2	PART B7
EVALUATOR 3	PART A3	PART A6	PART B3	PART B6
EVALUATOR 4	PART A4	PART A5	PART B4	PART B5
EVALUATOR 5	PART A1	PART A2	PART B1	PART B2
EVALUATOR 6	PART A3	PART A4	PART B3	PART B4
EVALUATOR 7	PART A5	PART A6	PART B5	PART B6
EVALUATOR 8	PART A7	PART A8	PART B7	PART B8
EVALUATOR 9	PART B3	PART A3	PART B1	PART A4
EVALUATOR 10	PART B2	PART A5	PART B4	PART A6
EVALUATOR 11	PART B6	PART A7	PART B8	PART A8
EVALUATOR 12	PART B5	PART A1	PART B7	PART A2

Table 2: Distribution of audio extracts among the evaluators

The evaluators were required to assess the intelligibility of each interpreting performance on a 6-point scale, where 1 was totally unintelligible and 6 perfectly intelligible, and the impact of pauses in the renditions, on a 5-point scale, with 5 indicating the most adverse impact. The results showed that the intelligibility score of both groups was identical, with an average score of 3.375 points out of 6, corresponding to a comprehensible rendition despite syntactic errors and poor lexical choices.

However, the analysis of the answers concerning pauses in the pronunciation of numbers in the TT showed a problematic factor related to the software's display of the numbers on the screen. The control group recorded an average score of 2.5 out of 5 points concerning the negative impact of pauses on the overall quality of interpretation, while the experimental group recorded a score of 3 out of 5 points. The types of pauses

were also different between the two groups, with the control group having a higher rate of long, empty pauses and the experimental group having a prevalence of short pauses[3].

The analysis of the intelligibility results seemed to suggest that the ASR system may have improved the naturalness in the delivery of the interpreting trainees and their confidence in returning the numbers in the target language.

4. Conclusions

The aim of this study was to analyse the impact of new technologies in the domain of conference interpreting, with a specific focus on the advantages and disadvantages of employing a CAI tool featuring integrated ASR software. The research sought to investigate the quality of output and handling of specific elements, such as numbers, and address the question of whether in-booth technological support tools could replace the work of the boothmate.

Preliminary analyses indicated that the software achieved an accuracy rate of approximately 80% in transcribing numbers.

With regard to the accuracy and completeness of the information content detected in the participants' renditions, the experimental group that received assistance from the software displayed slightly higher accuracy rates and fewer omissions compared to the control group. However, the latency of the software in transmitting transcriptions of numbers on the screen presented a problematic factor that reduced the possibility of using the hints provided while interpreting and affected the occurrence and distribution of omissions.

The quality of the TTs was then assessed in terms of intelligibility, which was found to be low for both groups. Various factors, including the high density of numbers in the ST, participants being students rather than expert users, human-machine interaction aspects, and tool usability, might contribute to this outcome. Nonetheless, it is important to note that the results suggest that external support, whether human or technological, cannot ensure precise interpreting or rendering of numbers.

As students and professionals alike become more interested in using computer assistance for interpreting, it is clear that becoming more familiar with these technological tools could lead to a shift in perception regarding their usefulness and potential benefits. To this end, it seems reasonable to suggest that a more comprehensive and focused training period could help current and future interpreters adjust their interpreting techniques to better handle software limitations, such as by developing specialised strategies to minimise the impact of software latency.

Despite the exploratory nature of our study and the presence of various methodological limitations and constraints, as mentioned above, we firmly believe that this subject deserves deeper investigation and could prove valuable for better integrating state-of-the-art CAI tools into the curriculum of trainee interpreters.

As a final reflection on the overall quality resulting from the twofold approach used to evaluate the TTs, the results of our study seem to suggest that replacing a

3 We opted not to provide guidelines to evaluators on distinguishing long and short pauses, wanting to capture their personal perception without overly binding criteria.

human boothmate with a digital boothmate would not lead to a significant decrease in the quality of a novice interpreter's rendition. In fact, it may even slightly improve it.

References

AIIC (1982) *Practical guide for professional interpreters*. Geneva: AIIC.

Baigorri-Jalón J. (2014) "From Paris to Nuremberg: the birth of conference interpreting", in H. Mikkelson / R. Jourdenais (eds) *The Routledge Handbook of Interpreting*, Routledge Handbooks Online (11-28).

Braun S. (2020) "Technology and interpreting", in M. O'Hagan (ed), *The Routledge Handbook of Translation and Technology,* London, Routledge, 271–288.

Braun S. / Clarici A. (1996) "Inaccuracy for numerals in simultaneous interpretation: neurolinguistic and neuropsychological perspectives", *The Interpreters' Newsletter* 7, 85–102.

Brunetti M. (2023) *Gestione dei numeri in interpretazione simultanea e spillover effect: proposta sperimentale per l'ampliamento della funzione di riconoscimento del parlato nei software di interpretazione assistita*, unpublished MA Thesis, UNINT, Università degli Studi Internazionali di Roma.

Canali S. (2019) *Technologie und Zahlen beim Simultandolmetschen: Utilizzo del riconoscimento vocale come supporto durante l'interpretazione simultanea dei numeri*, unpublished MA Thesis, UNINT, Università degli Studi Internazionali di Roma.

Carbutto M. (2022) *L'impatto delle nuove tecnologie per l'interpretazione simultanea nella resa dei numeri: un contributo sperimentale*, unpublished MA Thesis, UNINT, Università degli Studi Internazionali di Roma.

Carroll, J.B. (1966) "An Experiment in Evaluating the Quality of Translations". *Mechanical Translation* 9, 55-66.

Chevalier L. / Gile D. (2015) "Interpreting quality: a case study of spontaneous reactions". *FORUM* 13, 1-26.

Chmiel A. / Spinolo N. (2022) "Testing the impact of Remote Interpreting settings on interpreter experience and performance: methodological challenges inside the Virtual Booth", *Translation, Cognition & Behavior* 5/2, 250-274.

Collard C. (2019) *A Corpus-Based Study of Simultaneous Interpreting with Special Reference to Sex*. PhD dissertation, Ghent University.

Collard C. / Defrancq B. (2017) *Sex Differences in Simultaneous Interpreting: A Corpus-based Study*. Poster presented at the CIUTI Forum 2017, Geneva.

Defrancq B. / Fantinuoli C. (2020) "Automatic speech recognition in the booth. Assessment of system performance, interpreters' performances and interactions in the context of numbers", *Target* 33/1, 73-102.

Desmet B. / Vandierendonck M. / Defrancq B. (2018) "Simultaneous interpretation of numbers and the impact of technological support", in C. Fantinuoli (ed) *Interpreting and Technology*, Berlin, Language Science Press, 13-27.

Fantinuoli C. (2017) "Computer-assisted preparation in conference interpreting", *Translation & Interpreting* 9/ 2, 24-37.

Fantinuoli C. (2018a) “Computer-assisted interpreting: challenges and future perspectives”, in G. Corpas Pastor / I. Durán-Muñoz (eds), *Trends in E-Tools and Resources for Translators and Interpreters*, Leiden, Brill, 153-174.

Fantinuoli C. (2018b) “Interpreting and technology: the upcoming technological turn”, in C. Fantinuoli (ed) *Interpreting and Technology*, Berlin, Language Science Press, 1-12.

Fantinuoli C. / Marchesini G. / Landan D. / Horak L. (2022) *KUDO Interpreter Assist: Automated Real-time Support for Remote Interpretation,* Proceedings of the 43rd Conference Translating and the Computer, 68–77.

Fantinuoli C. / Montecchio M. (2023) “Defining maximum acceptable latency of AI-enhanced CAI tools”, in O. Vázquez / A. Correia / S. Araújo (eds) *Technological Innovation Put to the Service of Language Learning, Translation and Interpreting: Insights from Academic and Professional Contexts*. Berlin. Peter Lang, 213-225.

Frittella F. M. (2019) ““70.6 billion world citizens”: investigating the difficulty of interpreting numbers”, *Translation & Interpreting* 11/ 1, 79–99.

Frittella F. M. (2022) *ASR-CAI tool supported SI of numbers: Sit back, relax and enjoy interpreting?* Proceedings of the 43rd Conference Translating and the Computer, 88-102.

Frittella F. M. / Rodríguez S. (2022) “Putting SmartTerp to test”, *INContext: Studies in Translation and Interculturalism* 2, 137-166.

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

Goldsmith J. (2023) “Cymo Note: speech recognition meets automated note-taking”, *The Tool Box Journal*, 23-2-34, <https://www.internationalwriters.com/toolkit/current.html>.

Guo M. / Han L. / Anacleto M. T. (2023) “Computer-assisted interpreting tools: status quo and future trends”, *Theory and Practice in Language Studies* 13/1, 89-99

Kajzer-Wietrzny M. / Ivaska I. / Ferraresi A. (2021) “‘Lost’ in interpreting and ‘found’ in translation: using an intermodal, multidirectional parallel corpus to investigate the rendition of numbers”, *Perspectives* 29/4, 469-488, <https://doi.org/10.1080/0907676X.2020.1860097>.

Lamberger-Felber H. (2001) “Text-oriented research into interpreting: examples from a case-study”*, Hermes* 26, 39–63.

Mazza C. (2001) “Numbers in simultaneous interpretation”, *The Interpreters’ Newsletter* 11, 87–104.

Ordeig F. (2019) *A hands-on taste of RSI (Remote Simultaneous Interpreting): a brief personal view,* <https://aibarcelona.blogspot.com/2019/01/>.

Pinochi D. (2009) “Simultaneous interpretation of numbers: comparing German and English to Italian. An experimental study”, *The Interpreters’ Newsletter* 14, 33–57.

Pisani E. / Fantinuoli C. (2021) “Measuring the impact of automatic speech recognition on number rendition in simultaneous interpreting”*,* in C. Wang / B. Zheng (eds.)*, Empirical studies of translation and interpreting. The Post-Structuralist Approach*, London, Routledge, 181–197.

Prandi B. (2015) *Use of CAI tools in Interpreters' Training: A Pilot Study*, Proceedings of the 37th Conference Translating and the Computer, London, AsLing, 48-57.

Prandi B. (2018) "An exploratory study on CAI tools in simultaneous interpreting: theoretical framework and stimulus validation", in C. Fantinuoli (ed), *Interpreting and Technology*, Berlin, Language Science Press, 28-59.

Prandi B. (2022) *Computer-Assisted Simultaneous Interpreting. A Cognitive-Experimental Study on Terminology*. Berlin. Language Science Press.

Rodriguez S. / Gretter R. / Matassoni M. / Alonso A. / Corcho O. / Rico M. / Falavigna D. (2021) *SmarTerp: A CAI System to Support Simultaneous Interpreters in Real-Time*, Proceedings of the Translation and Interpreting Technology Online Conference, INCOMA Ltd, 102–109. < https://aclanthology.org/2021.triton-1.12>.

Rütten A. (2004) "Why and in what sense do conference interpreters need special software?", *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 3, 167-177.

Seeber Kilian G. (2015) "Simultaneous interpreting*"*, in H. Mikkelson / R. Jourdenais (eds) *The Routledge Handbook of Interpreting*, Routledge Handbooks Online, 79-95.

Setton R. / Dawrant A. (2016) *Conference Interpreting – A Complete Course,* Amsterdam/Philadelphia, John Benjamins.

Tiselius E. (2009) "Revisiting Carroll's scales" in C. Angelelli and H.E. Jacobson (eds.), *Testing and Assessment in Translation and Interpreting Studies*, Amsterdam/Philadelphia, John Benjamins, 95-121.

Good morning, everyone and thank you for being here. For my presentation, I will focus on the most important labour market indicators during the last years to give you an overlook on the EU employment rates, comparing pre-pandemic levels with the most recent data, based on surveys conducted in the first months of 2021.

Before the COVID-19 crisis hit, the EU was experiencing a steady, though decelerating employment growth. Continuing the positive performance in the labour markets that started in 2013, employment further expanded in 2019, reaching a record level at the end of the year. Reform efforts by Member States in the aftermath of the financial crisis contributed to this job-rich economic growth, though challenges persisted in some Member States and regions.

The COVID-19 pandemic reversed this trend, suddenly changing our ways of working and living. It has taken a significant toll in terms of human lives and caused an unprecedented economic shock, with great social impacts. The response by national authorities and European institutions has been swift. Member States have provided support to the groups and sectors that have been particularly affected. The implementation of these measures has, so far, avoided the emergence of a massive employment and social crisis across the EU. Yet, many uncertainties remain, especially about how long the pandemic will last, when a sustainable economic recovery will materialise, and about the consequences for people in vulnerable situations. The situation differs across countries, and Member States that already experienced serious socio-economic challenges before the pandemic are now even more exposed to vulnerabilities.

So, let's look at the situation in more detail.

Though signs of slowdown in employment growth were already evident in the second half of 2019, most labour market indicators have interrupted their positive trend at the outbreak of the pandemic. Total employment, which had increased by 15 million since mid-2013, fell by 6.2 million people between the fourth quarter of 2019 and the second quarter of 2020. This has been the sharpest decline ever observed over two successive quarters. It then increased slightly by 1.9 million in 2020.

According to Eurostat, the statistical office of the European Union, 13.374 million men and women in the EU, of whom 11.274 million in the euro area, were unemployed in March 2021. Compared with February 2021, the number of persons unemployed decreased by 85.000 in the EU and by 76.000 in the euro area. Compared with March 2020, unemployment decreased by 2.359 million in the EU and by 1.931 million in the euro area.

The swift adoption of short-time work schemes and other labour market retention measures, together with a decline in activity rates, have so far led to only a moderate increase in the unemployment rate. In this regard, youth unemployment, regarding people aged 15-24, increased more markedly than unemployment for other age groups, and the share of young people neither in employment, nor in education or training (NEET) soared.

In 2021, the youth unemployment ratio was 6.5%, regarding all people aged 15-24 who were unemployed. This share amounts to 16.6% if only the young people in the labour force are taken into account, the so-called youth unemployment rate. The youth unemployment ratio in the EU ranged from 2.2% in Czechia to 13.4% in Swe-

den, while the youth unemployment rate ranged from 6.9% in Germany to 35.5% in Greece.

Non-EU born workers have also been severely affected. These outcomes vary significantly across countries, regions and sectors and require close monitoring and policy efforts to avoid a more negative impact in the medium to long term. Before the crisis, this indicator, in the age group 20-64, rose steadily until 2019. It then dropped to 62.7% in 2020, corresponding to a decrease in absolute terms of 1.5 million persons, from 16.5 million in 2019 to 15 million in 2020, and almost 9% in relative terms.

As already mentioned before, massive use of short-time work schemes helped mitigate the consequences of the shock to the labour market, therefore the hours worked per worker fell abruptly, by 11.7% in 2020 compared to the last quarter of 2019. At the same time, absences from work increased significantly, from 9% in 2019 to 18.7% in 2020, and around half of this increase is due to temporary lay-offs.

Moving on to rates by sex, the gender gap in employment stagnated and the one in pay narrowed over the past five years, but the crisis has created new risks and underlined the need for reforms and investments. While female employment rates increased, the gender gap in pay has only improved slightly since 2013, in spite of the higher average educational attainment of women. In 2019, the employment rate of women has increased to 67.3%, 0.8 percentage points more than in the previous year. Indeed, women's over-representation in lower paid sectors and occupations, as well as in part-time employment, makes them particularly vulnerable in the labour markets struck by the COVID-19 crisis. Gender gaps are larger for women with young children: in 2019, they faced a negative employment gap of 14.3 percentage points in contrast to women without children, whereas men in the same situation saw a positive gap of 9.6 percentage points. Employment outcomes are poorer in particular for older, non-EU born and low-skilled women and women with disabilities. Female labour market participation could be strengthened by reforms and investments in early childhood education and care as well as long-term care services, and work-life balance policies.

Lastly, before the COVID-19 crisis, the number of people at risk of poverty or social exclusion was declining for the seventh consecutive year, though slow progress in countries with higher poverty rates hints at challenges for social protection systems. In 2019, more than 91.3 million people were at risk of poverty or social exclusion, which is 3.39 million fewer people compared to the previous year. These positive developments were nonetheless showing some deceleration in many Member States. The COVID-19 crisis, with the associated increase in unemployment and inactivity, makes the achievement of the Europe 2020 target of 20 million fewer people in poverty or social exclusion, compared to 2008, highly unlikely. The current situation poses increased challenges for social protection systems, in particular in relation to sustainably ensuring adequate incomes and the provision of quality services especially for all those who need them most.

If you'd like to know more about this, please visit our website where you can read everything about employment statistics in the EU and have a look at the data. Thank you for your attention.