universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Computer-assisted interpreting and automatic speech recognition: effects of the CASSIS software on simultaneous interpreting

verfasst von / submitted by

Benjamin Szilas, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 070 331 381
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Translation Deutsch Ungarisch
Betreut von / Supervisor:	Univ.-Prof. Dragos Ioan Ciobanu, PhD

Acknowledgements

I would like to express my deepest gratitude to Prof. Dragos Ioan Ciobanu for supervising this thesis. I could not have written this paper without his thoughtful feedback, and without his invaluable guidance. I would also like to thank members of my defence committee for their knowledgeable insights.

I want to sincerely thank Dr. Alex Hajós for allowing usage of the software tested in this paper. I hereby disclaim that no sponsorship was received from the software developers, nor did they influence the contents of this paper in any way, shape or form other than allowing access to the software for this study, for which I am very grateful.

I would like to thank all interpreters who participated in the study, devoting their free time, successfully completing the difficult interpretation task, and providing valuable feedback.

I would like to extend my utmost gratitude to my family and friends who stood by me and supported me throughout the writing process.

Table of contents

1. Abstract

Computer-Assisted Interpreting (CAI) tools are expected to see more widespread use in the future (Pym, 2011; Corpas Pastor, Costa and Durán-Muñoz, 2014; Corpas Pastor and Fern, 2016; Fantinuoli, 2018, Vogler, Stewart and Neubig, 2019). CAI tools that query glossaries supported by Automatic Speech Recognition (ASR) might aid terminological accuracy (Fantinuoli, 2017; Desmet, Vandierendonck and Defrancq, 2019; Prandi, 2019), but they may also "provide too much information and increase the cognitive load imposed upon the interpreter" (Stewart et al., 2018). Mellinger (2019) and Prandi (2019) identify a need to further investigate both the simultaneous interpreting (SI) product and cognitive processes when using CAI tools. Prandi suggests "future studies should also include the option to query the glossary through automatic speech recognition". This thesis will therefore investigate whether the CASSIS software can address some of these issues (Hajos, Hajos and Klein, 2020).

My first research question regarding this particular SI product is: How does this software influence terminological accuracy? To investigate this question, the methodology was based on Prandi's exploratory study design (2019). Participants interpreted a 10-minute-long English speech into German containing specialised terminology. 5 minutes were interpreted with CASSIS, 5 minutes without it. Terminological accuracy was determined according to the classification in Prandi's study design.

Secondly, the possible effects of terminological preparation on the results were investigated (Moser-Mercer, 1992; Will, 2007; Rodríguez and Schnell, 2009; Xu, 2018). Participants were asked to reflect on their terminological preparation prior to the study and how the CASSIS term base affected their terminological performance.

Lastly, implications of using CASSIS on cognitive processes in the booth were investigated, on the coping strategies used (Strolz, 1994; Gile, 2009; Seeber, 2011). The evaluated terminological accuracy of renditions was cross-referenced with a fatigue test of participants and a follow-up questionnaire (Ivanova, 2000) to gauge cognitive loads, problem triggers, and adaptation of coping strategies when using CASSIS.

Findings can indicate how CAI tools with automatic transcription and term base matches influence the terminological accuracy, and the interpreting process.

2. Literature review

2.1. Computer-assisted interpreting

2.1.1. Introduction

> Interpreting is a form of Translation in which a first and final rendition in another language is produced on the basis of a one-time presentation of an utterance in a source language. (Pöchhacker, 2016: 11)

Interpreting, regarded by some as one of the oldest professions, has been affected by the changes of the Information Age. At the advent of the Information Age, Moser-Mercer (1997) identified several key areas when researching the effect of working with computers in interpreting: “integration of information across sensory channels, multiple visual information sources and their impact on attentional control and memory, (…) stress-related issues of working with computers and video images, and many more” (ibid.: 195)

Technology has thoroughly reshaped the world of Translation in general over the past decades, especially because computers are particularly useful for storing and searching through thousands of pages of information in an instant (see Pym, 2011). He observes that the use of technology transforms how products in the translation profession are created. For example, translators using translation memories, or post-editing machine translation outputs have largely different workflows than translators before the Digital Age. (Pym, 2011: 2)

He envisions that an incoming or ongoing shift towards a more technologized interpreting profession will bring forward similar shifts in workflow, and even shifts in the professional community. While it is difficult to foresee such future changes, the workflow of interpreters has certainly been changed by the Digital Age. A few decades ago, when comparing interpreters’ and translators’ workflows, Barbara Moser-Mercer wrote:

> Interpreters – quite used to more “visibility” than their colleagues in translation – have by and large not benefitted from these technological advances, perhaps mostly because available products have not really met their needs, and in part because they

represent a rather small market segment. (Moser-Mercer, 1992: 507)

Using computers and the Internet for obvious communication purposes is hardly a novelty anymore and there have been software developments for specific use by interpreters (Pöchhacker, 2016: 184), yet scholars often speak of an apparent disinterest (Pym, 2011; Corpas Pastor, Costa and Durán-Muñoz, 2014) and even a "history of denial" (Ortiz and Cavallo, 2018: 11). There is an urgent need for more research (Fantinuoli, 2019) when it comes to computer-assisted interpreting (CAI) software or use of interpreting-specific information and communication technology in general.

Data about this disinterest can be gathered from the low return rate of a survey about "technology tools for interpreters" (Corpas Pastor and Fern, 2016). The survey was sent out to 47 international translator and interpreter groups and associations and distributed in online forums and groups (ibid.: 40-42). Of these groups, the ATA (American Translators Association) alone has about 10,000 members. Out of the tens of thousands of people whom this survey had potentially reached, 133 have completed it, which corresponds to a rough estimate of less than 1% return rate, much lower than what is deemed a good survey return rate (Dillman, Smyth and Christian, 2014). Corpas Pastor and Fern hypothesise that the surveyed tools are unsatisfactory for interpreters as study respondents expressed e.g., "sometimes our minds work faster than a computer" (ibid.: 35). The types of technology used by the surveyed interpreters either in the preparation phase or during an assignment were the following: audio recordings, computer aided translation tools, translation memory systems, machine translation systems, term extractors, term banks, concordancers, corpora, web-based resources, bilingual or monolingual dictionaries/glossaries/thesauri, e-journals and e-books. (ibid.: 54)

2.1.2. Definition of computer-assisted interpreting

The term computer-assisted interpreting (CAI) was coined analogous to computer-assisted or computer-aided translation (CAT). However, while CAT tools have been widely used in the translation industry for decades, CAI tools are a newer advance and are, as discussed based on Corpas Pastor (2015) and Pym (2011), not as widespread yet. Therefore, finding already existing definitions for CAI can be a challenge. Moreover, the term computer-assisted interpreting has not yet found

general acceptance and usage yet, as there are a multitude of term used for this phenomenon, with sometimes differing meaning. This chapter will therefore attempt to categorise some of the terms in literature that are somewhat analogous to CAI and choose a definition that is suited for this research. While an attempt to create a complete overview of all technology for interpreters is beyond the scope of this thesis, this chapter will define terms for the purpose if this paper and attempt to place and define CAI tools within the realm of interpreting technology.

As an example for ambiguity, the 2019:20539 ISO standard for translation, interpreting and related technology, unsurprisingly, includes definitions for CAT tools and TM systems under *translation technology*, however is largely limited to describing booths and other audio equipment when it comes to *interpretation technology* (ISO, 2019). Other papers include another, wider scope of available technology for interpreters also including hardware (e.g. smart pens for consecutive note taking) and thus speak of the wider term *technology-assisted interpreting* or technology tools (see Costa, Corpas Pastor and Durán-Muñoz, 2014; Corpas Pastor and Fern, 2016). The term *technology tools* however can be ambiguous in this context, as technology could also mean equipment that enables interpretation modes such as simultaneous interpreting (SI) booths or software solutions for remote interpreting (RI), not to mention automatic speech-to-speech translation, alongside respeaking (Pöchhacker, 2016: 182–190). Due to all these available technologies, I will refrain from using the term *technology tools* or *technology-assisted interpreting* in this paper when talking about CAI tools and elaborate on a more distinct terminology.

Fantinuoli mentions that CAI tools can be considered the interpreter's counterpart of CAT tools, "both having an influence on the translation process and product, on the workflow, etc."(2018: 156) In recent years, many publications (Tripepi Winteringham, 2010; Fantinuoli, 2017; Prandi, 2017; Fantinuoli, 2018; Stewart *et al.*, 2018; Fantinuoli, 2019b; Prandi, 2019; Mellinger, 2019; Vogler, Stewart and Neubig, 2019) have adopted the term CAI. Other papers have used the term "specialised software" (Pöchhacker, 2016: 185) when talking about software aiding the interpreting process.

As an overarching term, some papers use "Information and Communication Technologies (ICTs) – that is computers, digital tools and the Internet" (see Tripepi

Winteringham, 2010: 87). The term ICTs also includes cases when software or websites, which are not specialized for interpreting, are used to aid an interpreting task. It is worth noting, however, that computer-assisted terminological research can be practised without specialized software or hardware, since the "World Wide Web with its unprecedented richness of subject and terminological information, for example, has changed the way interpreters prepare their assignments" (Fantinuoli, 2018: 154). One could argue that this is also computer-assisted interpreting, since computers aid the preparation phase of an assignment. ICTs are also used to assist interpreter training from digitalization of material to simulated conferences in a virtual reality environment (Sandrelli, 2015).

2.1.3. Distinction between hardware and software

Fantinuoli uses the terms "process-oriented on the one hand, and the setting-oriented technologies on the other" to differentiate between ICTs (2018: 155). Process-oriented technologies, he argues, aid the interpreting process in the preparation or the actual interpretation phase, while the setting-oriented category includes "ICT tools and software 'surrounding' the interpreting process proper, such as booth consoles, remote interpreting devices, training platforms, etc.". However, he agrees that these two categories are not clear-cut, as "for example, it is not easy to decide which of the two categories should be addressed by the Consecutive Pen" (ibid.: 156). The Consecutive Pen

> "consist of a microphone, a built-in speaker, 3D recording headsets, and an infrared camera. They are used to take notes – they have a normal ink cartridge and are held as 'normal' pens – and to capture data on a microchipped paper. (…) At any time, (…) the user of the pen can play back the speech from the notes taken on paper, allows users to backup, search, and replay notes from their computer. Users can also upload and convert notes to interactive Flash movies or PDF files." (Orlando, 2010: 77–78)

One amendment to this categorisation into setting and process-oriented could potentially eliminate some ambiguities: to make a clear distinction between *hardware* and *software*. Both terms are unambiguous in our everyday lives: hardware are physical components, which are used to run software on them. These two categories have implications for the interpretation process as well. Hardware is always setting-

oriented in the sense of Fantinuoli's paper, "as they primarily influence the external conditions in which interpreting is performed or learned" (2018: 155). While none would see much use without the other, both have different implications for the profession. Laptops, tablets, etc. must be physically present all the time and thus require spatial attention from interpreters. Judging by the overall hardware trend that we can see in our everyday lives, interpreters might also favour ever smaller, lighter hardware, as long as they keep enough processing power to run desired software. Recent developments such as the Smartpen (Orlando, 2010, 2014) show the possibilities when keeping in mind the above trends for hardware as the pen is light, easy to carry, yet is still able to run software that records, saves and plays back audio and recorded notes as well.

However, this paper will focus on software made for interpreters, as it seems to have wide possibilities open for scientific investigation. Furthermore, CAI software

> appears an underrepresented subject within interpreting studies in general, and technology related studies, in particular. The fact that interpreters increasingly rely on software, both interpreter-specific and not, to support their daily professional life (…) makes new technology in interpreting an interesting research subject which requires to be analysed in detail. (Fantinuoli, 2018: 156)

Software used by interpreters can be either setting-oriented, such as videoconference software upon which the act of remote interpreting relies, or process-oriented, "such as terminology management systems, knowledge extraction software, corpus analysis tools and the like". However, any classification can have blurred boundaries between the categories. Software such as InterpretBank will be used to look up terminology while interpreting, thus the interpreter is to some degree reliant on it to deliver the interpretation. One could argue that this is makes it both setting-oriented and process-oriented as well. (ibid.: 155)

For this paper, *CAI* will be defined according to Fantinuoli (2018: 155) as *software* used by a *human* interpreter *designed to aid* them at some stage through the interpreting *process*, whether it happens during preparation and debrief or during interpreting. This differentiates them from *setting-oriented software*, which are used to *enable* some form of interpreting setting or mode altogether. For example, videoconference interpreting would not be possible at all without the bespoke setting-

oriented software. Therefore, videoconferencing software creates the foundations for the interpreting setting.

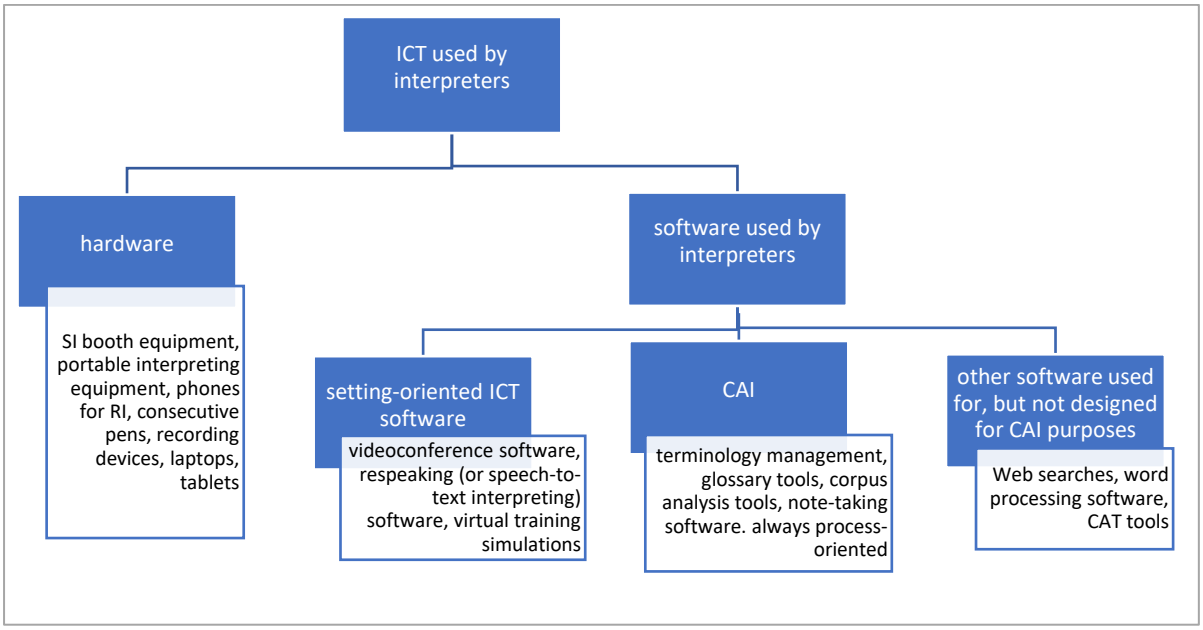


Figure 1. CAI in the realm of information and communication technology used by interpreters

2.1.4. The need for CAI terminology tools

Given the role that computers play in storing and accessing information in an instant, it is perhaps self-explanatory that many CAI tools aim at extracting and managing terminology. Ortiz and Carvalho enlist CAI tools according to their main purpose, considering which of these tools are being updated at the time of the publication (Ortiz and Carvalho, 2018: 17-19). Will (2015) aims at developing a "series of transparent and objective criteria for the use and the conception of the relevant software" (ibid.: 179). After evaluating the software *interplex, Terminus, LookUp Prof.*, and *InterpretBank 3* in their respectively available versions to the date of the paper, he calls for "the development of a suitable and powerful expert system for conference interpreters" (ibid.: 179)

Interpreters acquire information about the event beforehand, in the *preparation phase* of an assignment. They build up knowledge about the specialised area and the needed terminology and they can rely on this background knowledge while interpreting to understand the source text and find the right terms in the target language. (ibid.: 182) Will therefore argues that glossaries containing only SL (source language) and TL (target language) term pairs, which are frequently used by

interpreters, are insufficient. He lists additional information, or *metadata*, that helps build the knowledge system (ibid.: 184). These are on the micro level: synonyms, abbreviations, contexts, and degree of equivalence in both source and target language. Information on the macro level can include a definition and relation to other specialised terms. Additional information can include the client, conference name and date, conference topic, entry date, author, and source(s).

Will argues that a terminology CAI tool needs to display this information differently during the *interpretation phase*. (ibid.: 186) During SI, the source text is received in a spontaneous and linear manner, independently of the interpreter, thus it is not possible to 'pause' source text presentation to research or look up previously researched terminology. The time that passes between interpreters hearing a source text term and them uttering the target language equivalent is the *ear-voice span* of that term (ibid.: 181) He postulates that a terminology tool for SI thus needs to present terminology entries during the interpretation phase in such a manner that they can be used adequately within the ear-voice span (reduce the amount of additional information displayed according to interpreter preference) furthermore these software need an adequate search and filter function to use during interpretation. (ibid.: 186)

Will sums up three features which can put a terminology CAI tool above paper glossaries and Excel sheets in terms of functionality (ibid.: 186-187).

-flexible view, meaning the ability to switch between a 'preparation view' and a 'conference view' and to customise these according to preference.

-database and help functions: different ways to search and retrieve information from the database; short-stroke, error-tolerant, and diacritics-independent search results. Automatised collection of metadata such as date of entry and last modification.

-ergonomic use: intuitive functionalities, tutorials, compatibility with as many data formats as possible.

This is a prescriptive list of functionalities based on theory. It is not mentioned if and how it is possible to implement all these features in a working software, or if interpreters would indeed use these features. Therefore, the above recommendations will be cross-referenced with a study by Barbara Moser-Mercer (1992). She surveyed the features which interpreters look for in a CAI terminology management software.

On the micro level, 75% or more interpreters expressed the wish to enter the following *metadata* in a terminology software: TL and SL term, definition, synonyms of the SL term, reliability of the TL term, and subject field (ibid.: 521). 80% or more wished for *database and help functions* that make different types of term retrievals possible: alphabetical and subject-oriented print-out, alphabetical look-up, parallel print-out of SL and TL term pairs. However only 50% requested a short-stroke lookup and only 30% requested support for non-Latin characters. The latter question is now obsolete as the UNICODE standard supporting more than 100 scripts and alphabets has been widely accepted since then. Apart from the short-stroke lookup, Moser-Mercer's results can indicate what specific functions interpreters find useful in a terminology management software, although an up-to-date verification could yield new results.

One could also question the need for developing software to suit SI, given that the task can be done without using terminology software, by simply relying on memory, and printed glossaries.

However, there are situations where CAI software with the functions proposed by Will can be of use during SI. For example, what if there is not enough time in the preparation phase to memorise all necessary terms? Of course, one can bring a paper glossary, but the more terms are on the paper, the harder it will be to pinpoint a specific term during SI. Search functions have the potential to scan through more term entries in a shorter time than the human eye. There is also a benefit of accessing previously worked out terminology: terminology software can store term entries virtually indefinitely, which can then be accessed, and updated at the touch of a few buttons even years after previous use.

However, Corpas Pastor, Costa and Durán-Muñoz found in their 2014 paper that

> most interpreters seem to be unaware of the opportunities offered by language technologies. As far as terminology is concerned, interpreters continue to store information and terminology on scraps of paper or Excel spreadsheets, while the use of technologies and terminology management tools is still very low (ibid.: 68).

Similarly to Will (2015), Corpas Pastor, Costa and Durán-Muñoz, (2014) are investigating functionalities of terminology management software. They argue that “interpreters require specific tools to meet their needs, which are different from

translators and terminologists" (ibid.: 69). However, the notion that interpreters use terminology management tools differently than translators is not reflected in their evaluation, as their given scores are based on the amount of available features of said tools (see ibid.: 74), regardless of whether those features are used or needed by interpreters.

Studies by Prandi (2017, 2019) aim at creating a framework to investigate the usability of CAI tools, terminological quality, and implications for cognitive processes when using CAI tools. Her studies evaluated interpretations of 3 speeches with a speed of 122 words per minute, each 12 minute long and containing specialised vocabulary. The speeches were recorded by an English native speaker and interpreted into German. One of the speeches was interpreted using a Word glossary, the second an Excel sheet, the third using an InterpretBank glossary. The speeches were purpose-written for the study, with each of the three containing 36 specialised terms. She recorded how many glossary searches were made, whether the searched terms were in the middle or at the end of sentences, and usage of alternative coping strategies. She also evaluated the renditions according to three error categories: *close rendition*, *acceptable rendition*, and *zero/unacceptable rendition*. (Prandi, 2019: 41f)

Regarding glossary usage, she found that 5 participants looked up between 50-76% of the terms in the glossary with similar frequency. One participant searched significantly less, only 30% of terms in the glossary. 4 out of the 6 participants produced more close renditions in their interpretations with InterpretBank than using Excel or Word glossaries. 2 out of the 6 performed worse with InterpretBank in the same regard than with Excel or Word. Prandi concludes that

> 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. (Prandi, 2019: 56)

There are three more points which make further investigation necessary. **First**, the small sample size makes it difficult to reach conclusions, yet one third of participants performed worst with InterpretBank which cannot be glossed over. **Second**, it was not considered whether mental fatigue could have played a role in the

study results. Since each participant interpreted a total of 36 minutes, it would be worth investigating how that affected their cognitive capacity. In case a participant has been affected by mental fatigue towards the end of the 36-minute interpretation, their rendition fidelity may also have been affected by that. Results of the glossary type they have used last could be affected more than the results with the other two glossary types. It is furthermore unclear whether they have taken any breaks and in what order they were using each glossary. **Third**, the recent advances in automatic speech recognition technology, that were not tested yet within Prandi's study, create the need to evaluate CAI terminology tools with integrated ASR.

2.1.5. Surveys on CAI terminology tools

Another step into evaluating the usefulness of CAI tools are surveys of user satisfaction and studying user's feedback and habits. Barbara Moser-Mercer (1992) surveyed some features that interpreters would like to have in a CAI terminology management software, to better suit their terminological workflow. More than a third (37,7%) of the 122 respondents to her survey did not use or have access to a computer, possibly because "most international organisations are not nearly as well equipped electronically as the private market tends to be" (ibid.: 511).

Fast forward two decades, Corpas Pastor and Fern still suggest that interpreters seem to benefit little from CAI tools: while there is a "vast software industry which addresses the needs of professional translators in most fields", their assessment is that "in the world of interpreting, language tools are very poor and unsatisfactory" (2016: 12-13). However, they proceed to list terminology software for interpreters and their main available functions, without arguing why they would be very poor.

Similarly, Will also suggests that "there are only a few terminology management programs available for the specific needs of conference interpreting" (2015: 179). A paper by Costa, Corpas Pastor and Durán-Muñoz lists requirements for interpreter's terminology management software:

> Due to their working conditions, translators usually prefer to consult multiple definitions and contexts to find the best solution for the translation problem. On the contrary, interpreters will rarely have the time to go over multiple definitions, contexts, etc. to find the right one, and thus, they will need to store the most concise information

> to be able to consult it in the quickest and easiest way. Their responses in this survey also showed that the way they retrieve terminological information was context-specific, and that there was also a significant variation among individual interpreters. Flexibility is, therefore, of great importance to interpreters due to the variation of their context-specific terminology management practices, and on their individual preferences regarding the storage, organisation and retrieval of terminological information (Costa, Corpas Pastor and Durán-Muñoz, 2014: 69)

Gloria Corpas Pastor and Lily May Fern (2016) have surveyed 133 professional interpreters across all interpreting modalities "whether they use technology tools and resources during an interpretation", and "48.12% answered yes". Even though the return rate in their study was low, this is perhaps the most up-to-date data there is about usage of CAI tools. Although the authors did not specify further what kind of technology tools they mean, which makes this number difficult to interpret – were respondents to the above question only thinking of computer-assisted interpreting or other technology used by interpreters (e.g., laptop computers?)

Perhaps more telling is their graph depicting the usage of 15 different types of "technology tools" prior, and during an assignment. While the web-based technologies were used in 78 out of 133 cases during preparation, bespoke CAI software, for example term extractors were only used in 15, and term banks in 31 cases. 22 people reported using CAT tools - i.e., translation memory (TM) tools - to prepare for interpreting assignments. During the interpreting phase of an assignment, *web-based resources* represent the second most widely used out of 15 different tools, only to be surpassed by *bilingual dictionaries/glossaries/thesauri*.

Noteworthily, the used categories in this study are unspecific: "web-based technologies" could mean anything from a Google search, an online forum, or a specific corpus management and terminology research tool like the Sketch Engine. While a paper dictionary may not be practical to consult during interpretation due to the sheer time needed to look up words, glossaries may be less time-consuming depending on their size. "Bilingual dictionaries, glossaries and thesauri" can be found both in paper form and in web form, which is not specified either. Provided that the latter is the case, this data would suggest that while using computers (or devices connected to the internet) for interpreting assignments is not so rare anymore, bespoke CAI tools are used less often. The authors recorded that many interpreters

“simply do not have enough time during an interpreting task to consult technology tools” (ibid.: 35). Therefore, the study authors argue that “the next step is to create a tool that can respond immediately to a terminological problem”.

2.1.6. CAI with automatic speech recognition

Automatic speech recognition technology has been proposed by Fantinuoli (2017) to automate glossary searches during interpretation that were previously done manually.

> The main drawback is that the database is queried manually, adding an additional cognitive effort to the interpreting process. This disadvantage could be addressed by automating the querying system through the use of Automatic Speech Recognition (ASR), as recent advances in Artificial Intelligence have considerably increased the quality of this technology (Fantinuoli, 2017: 25)

The software proposed by Fantinuoli is a CAI tool that aims at eliminating the time needed to manually search for glossary entries during interpretation. The glossary is maintained by the interpreter in the InterpretBank software, during the preparation and debrief phases of interpreting. During interpretation, the ASR engine monitors the spoken source input and shows any matches within the prepared terminology databank. Moreover, it shows numbers as well, which have been identified as a common problem trigger in interpretation (Gile, 1995/2009). A trial version is available online for this software, with which the number recognition part can be tested for a limited amount of time.

Since the proposal of ASR integrated in the tool InterpretBank, there have been studies investigating the use of ASR in interpreting (Ortiz and Cavallo, 2018; Desmet, Vandierendonck and Defrancq, 2019), albeit the former focuses on theoretical gains using ASR, the latter uses a mock-up simulation to gain first conclusions. Ortiz and Carvallo point to the possibility of such a system helping interpreters but argue its limitations, namely a “*lack of an ASR package or solution that is economically accessible, can be used and installed on a computer (there are online solutions, but they may incur in confidentiality violations), performs well (processing speed), and is multilingual.”* (ibid.: 25f) Desmet et. al. investigated whether “*displaying numbers on a screen in the conference room, immediately after they have been articulated by the speaker, increases the accuracy of numbers in the target text*.”

Their experimental setup simulated an ASR software that is transcribing numbers in real time as the source speech is being interpreted. They found that the ASR simulations have improved the accuracy of all numbers (whole numbers, complex numbers, decimal numbers, and dates) in the interpreters' renditions. It reduced the total amount of omissions from 106 to 39, and the number of approximations from 39 to 4. They stress however, that they simulated a perfect transcription of numbers, whereas a working ASR software "*might not achieve perfect recognition and minimal latency*" (ibid.: 25).

2.1.7. The CASSIS web-based tool

Another CAI glossary tool with ASR has been put forward by Hajos, Hajos, and Klein (2020) and made available for the purpose of this study. In this web-based software named CASSIS the integrated ASR transcribes the entire source text in real time on screen. It uses the Google speech-to-text service and supports 20 input languages: Arabic, Bulgarian, Croatian, Czech, Danish, English, French, German, Hungarian, Japanese, Mandarin Chinese, Norwegian, Polish, Portuguese, Romanian, Russian, Slovak, Slovenian, Spanish, and Swedish. Numerals are only shown as part of the full transcription and not shown separately. Matches in the previously uploaded glossaries are highlighted in the live transcription and showed separately along with their target language pairs. This software design can have implications on the interpreting process, which will be described in the following chapters. Albeit the CASSIS software only supports bilingual word lists with the source language term on the left and the target language term on the right, this format corresponds to the reported usual practice of interpreters when preparing glossaries for the interpretation phase. (Moser-Mercer, 1992) In order to use a word list, CASSIS draws on glossaries from the memoQ CAT tool, or on bilingual glossaries exported in CSV format from a software of choice. This means that CASSIS does not have terminology management functions on its own, and its user relies on using another software for that purpose. This is one of the differences to InterpretBank, which offers terminology management functions as well. Another important point is that the ASR in CASSIS is cloud-based, which has its confidentiality implications (Ortiz and Cavallo, 2018: 26) that have to be taken into account by interpreters using it. The transcription in CASSIS is thus handled by a remote server rather than by the laptop or PC of the interpreter. Therefore, the internet connection has to be encrypted according to the confidentiality of the

interpreted material. The CASSIS software is also web based, that means there is no need to install any software. Accessing the CASSIS website with login information is enough to run the software and to upload new glossaries for the interpretation. Below is an illustration of CASSIS with buttons on top, transcription on the left, and term base matches on the right.

Start/stop ASR	Choose input language	Choose font size	Upload term base
Transcription		Term base matches	
This is where the transcription will appear. You can adjust the font size for readability. Any terms which are in the term base are highlighted. Their target language pair appears on the right side.		term → Begriff, der	

Figure 2. Illustration of the ASR transcription and automatic matches from the term base

2.1.8. Accuracy of automatic speech recognition software

> Automatic speech recognition (ASR) is an independent, machine-based process of decoding and transcribing oral speech. A typical ASR system receives acoustic input from a speaker through a microphone, analyzes it using some pattern, model, or algorithm, and produces an output, usually in the form of a text (Lai, Karat and Yankelovich, 2008).

ASR technology has reached the *word error rate* (WER) of roughly 5 %, and in case of transcribing numbers, close to 0% measured on the Dragon NaturallySpeaking system by Fantinuoli (2017: 32). These numbers refer to English as input language. Not all languages are supported by ASR technology to the same extent and other input languages may result in worse error rates. Ortiz and Carvalho (2018: 23) note that a 5% WER is "the same error rate humans experience in the comprehension of words during a conversation, estimated by IBM at 5.1%". One could argue that in the case of a transcription, even a 5% WER is significant and can alter the meaning of transcribed sentences. To illustrate that, here are some examples of errors made by ASR systems (Levis and Suvorov, 2012: 3f). Misrecognising (e.g. '*sinking'* as '*thinking'*) merging (e.g. '*deep end'* as '*depend'*), and splitting words (e.g. '*euthanasia'* as '*youth in Asia'*). While an attentive listener can recognise which one of

the possible versions is misheard and which one fits the context or the accent of the speaker, ASR systems will write the misheard versions out. The reader of the transcription will therefore have to realise the mistake and guess what the ASR system has "misheard", and the affected sentence may possibly be nonsense due to the mistake. This aspect therefore also has to be taken into consideration when evaluating the CASSIS software.

The availability of *speaker-independent* systems has been another significant step in the development of ASR (Fantinuoli, 2017: 28). A *speaker-independent* ASR system's "*training database contains numerous speech examples from different speakers so the system can accurately recognize any new speaker*" (Levis and Suvorov, 2012: 3) That being said, ASR systems are being trained mainly on native speaker accents, thus there can be a loss of quality when transcribing non-native accents. Derwing, Munro and Carbonaro (2000) measured Dragon NaturallySpeaking which has reached 10% WER as advertised in the version available in the year 2000 on sentences with a native English accent. Sentences uttered by non-native English speakers however, had a much higher, around 27-29% WER even though "Cantonese and Spanish speakers who participated in this study were all highly proficient, educated speakers of English" (ibid.: 600)

Even the performance of modern ASR systems utilising neural networks depends heavily on the accent of the speaker, be it native or non-native. Viglino, Motlicek and Cernak based a 2019 experiment on the Deepspeech 2 ASR system trained on different US, UK, Australian, Canadian, and Irish accents. The system performed a baseline value of 23,3%[1] with the pre-trained accents. They measured an 1,6% increase in WER to 24,9% when transcribing English with a New Zealand accent that it was not trained for. They decreased this difference slightly to 0,4% (15,1% compared to 15,5%) by making the software carry out an accent classification task simultaneously to the ASR task. They "assume that this is due to the presence of Australian in the training set" (ibid., 2019), which is somewhat similar to the tested New Zealand accent. The same process was repeated with an English input with an Indian accent. They initially measured a 31,9% WER increase (from 23,3%) to in total

[1] This WER is much higher than the recommended 5%. However, this study tested the Deepspeech ASR largely on monograms and bigrams, which prevents the software from scanning for context. WER in a continuous text would likely be lower.

55,2% compared to the control group values. The accent classification algorithm resulted in an even bigger gap to the control material, a 35,1% WER increase (from 15,1%) to a total of 50,2% in the transcribed Indian English accent.

This goes to show that ASR of today's standard will also be heavily affected by the accent(s) inherent in its training material. Accents not included in the development of an ASR system will likely be transcribed at reduced accuracy. This must be considered by interpreters using CAI tools with ASR. For example, if interpreting for a non-native English speaker, usability of an ASR system might be significantly limited.

2.2. Terminological preparation of interpreters

2.2.1. Moser-Mercers terminological preparation model

In order to shed light on how CAI tools can be shaped for the needs of interpreters, this paper will review some of the current models proposed for the terminological preparation process of interpreters. Terminology, as shaped by the works of Eugen Wüster "is centred around the concept of the '*term'* which he describes as a dual entity, consisting of a word form (*denomination*) and a content (*concept*) or meaning." (Will, 2007: 2) This duality of denomination and concept is incorporated into terminology-oriented works in translation studies (ibid.: 2-3) "Because terminology deals with specialized domains of knowledge, terms refer to the discrete conceptual entities, properties, activities or relations that constitute the knowledge of a particular domain." (Bowker, 2019) Thus, terms gain their unique meaning through their usage within a particular domain, which means that talking of terminology is connected to specialised domains or fields of knowledge.

Moser-Mercer (1992) has observed differences in terminological work of (simultaneous) interpreters compared to that of translators. Interpreters usually prepare for an assignment on shorter notice. The result of their terminology work is noted down differently to written translators: "interpreters will usually establish terminology lists with source language terms down the left margin and the centre of the page and target language terms down the right margin" (Moser-Mercer, 1992: 509) This is a simpler format than translator's terminology databases. She adds that need

for terminological research for conference interpreters may arise throughout the entire conference as “further information is added: new terms are acquired, etc.” (1992: 509). Highlighting the contrast to terminological work of interpreters vs. translators, she found that documents specific to the respective conference are most useful for the surveyed conference interpreters, while external terminological databases were perceived by respondents to be the least used terminological resource. These findings would potentially need to be updated, due to new ICT solutions suited for interpreter’s terminology research and extended to preparation for interpreting settings other than conferences. Summing up her survey, she highlights that software tailored to the terminological workflow of conference interpreters would enable them to be “more economical in dealing with increasing amounts of information and thus ultimately provide a better service”. (ibid.: 512)

2.2.2. Rodriguez and Schnell’s terminological preparation model

Rodríguez and Schnell (2009) stress that a conference interpreter’s terminological work is not only a preparation, but rather a process that goes on throughout the entire conference. New terms are researched and learned in three phases:

> first, during preparation for conferences, when the interpreter learns and memorizes the terminology; second, during work in the booth, where it is sometimes necessary to search for terminology because of the conditions inherent in the occupation, such as random or late sending of conference documents that are often incomplete and rarely multilingual; and, third, after work in the booth, when the interpreter follows up on the terminology work done beforehand (Rodríguez and Schnell, 2009)

While it is unclear from their paper, how often terminology work is carried out by one of the persons in the booth, it is important to clarify that research cannot be done by the person interpreting (see Will, 2007: 3). Researching a term in the booth would understandably put a high cognitive strain on boothmates as well, as they pay attention to the conference, although not actively interpreting it. One can easily see the problem in this case: how can a previously unresearched term be researched and memorised in such a scenario? CAI tools, with their possibility to quickly search through term bases can save valuable time and effort during such terminology work. Software with automatic speech recognition, as described in Chapter 3.5., can relieve

the interpreter from needing to memorise a previously unresearched term, or visually searching it on written down notes provided the quality of the ASR system is suitable.

2.2.3. Will's terminological preparation model

Will elaborates on these ideas and details a model of interpreter terminological preparation for assignments. He describes three phases of terminological preparation (Will 2007: 6-7) In *Phase I*, interpreters receive information and documents from their clients about the assignment. Research is conducted and the interpreter aims at "systematic and holistic knowledge acquisition geared towards the anticipated needs of the ensuing conference" (ibid:6). Terms are researched based on the anticipated background knowledge system of conference participants.

Phase II happens during the conference, as the interpreter encounters "previously unidentified knowledge" and terms. At the conference however, conditions for terminological preparations deteriorate, and researching new knowledge becomes limited to individual terms. In the booth, only term retrieval is possible (see Will, 2007: 7). "While both translator and interpreter have access to a wide variety of resources, the interpreter's access to these materials becomes extremely limited once he (sic) is in the booth" (Moser-Mercer 1992: 509)

Phase III is a debrief that consists of revising, re-structuring terminology collected during phase II, and incorporating it into the interpreter's individual terminology database.

2.2.4. Implications of CAI tools for terminological work

CAI tools can thus aid phase II of terminological work, as they can store terms which are researched and added during the conference but cannot be memorised due to the short time left before interpreting. CAI tools for terminology management can be applied during this stage, to aid the recall of information from memory. Importantly, however, this automatic term retrieval does not replace a thorough preparation on the needed background knowledge in the above-mentioned phase I. Interpreters obviously need to understand the underlying *terminological concepts* to meaningfully put the automatically retrieved *denominations* into context and into the target text. However, Xu (2018) argues that a successful preparation in phase I "does not

guarantee that terms are automatically ready for active use by the interpreter (especially in the simultaneous mode) (ibid.: 33)". A cognitive overload (Chapter 2.3.3.) could prevent recollection of terms from memory. CAI tools with automatic speech recognition can therefore serve as memory triggers, especially because as discussed, terminology work at conferences can continue up to the last second making memorising difficult.

CAI terminological tools supporting the first phase of terminological preparation have been evaluated (Corpas Pastor, Costa and Durán-Muñoz, 2014; Will, 2015), as well as their effectiveness studied (Prandi, 2019). However, some have raised concerns that these computer programs could potentially "provide too much information and increase the cognitive load imposed upon the interpreter" (Stewart et al., 2018). Therefore, there is a need for further research to investigate if and how CAI tools with ASR features affect the underlying cognitive processes of interpretation (Mellinger, 2019). First, an overview of cognitive processes in the booth will be provided and the effect of terminological software with ASR discussed.

2.2.5. Terminological focus

Terminological accuracy, which is measured in this study, is only one parameter of the interpretation product. Bühler (1986) studied what qualities interpreters look for when evaluating the performance of their peers. Criteria regarding the interpretation product were *native accent, pleasant voice, fluency, logical cohesion, sense consistency, completeness, correct usage of grammar, correct terminology,* and *appropriate style.*[2] Most important according to this study is *sense consistency* with the original message; however, the use of *correct terminology* was also considered important or highly important by all participants. This survey was later replicated (Kurz, 1993, 2001) with similar parameters, but asking conference participants instead of interpreters:

> Where was fairly high agreement by all groups on the importance of some criteria (sense consistency, logical cohesion, correct terminology), conference interpreters and users as well as different user groups among themselves differed in their assessment of the

[2] She also studied other criteria regarding the workflow and personality of interpreters: *thorough preparation of conference documents, endurance, poise, pleasant appearance, reliability, ability to work in a team, positive feedback from delegates.* Those may influence the interpretation, or the status of interpreters, but are not qualities of the interpretation product itself.

importance of other criteria (correct grammar, pleasant voice, native accent) (ibid. 2001: 398)

This shows that the correct use of terminology, although important both for interpreters and users, is only one aspect of interpretation products.

2.3. Cognitive models in interpretation

2.3.1. The cognitive level of interpreting research

When summarising literature about CAI technologies, Mellinger (2019) observes that despite a growing number of papers on the topic, “one area yet to be explored in great depth is the intersection of technology and interpreter cognition.” (ibid.: 34) Cognitive models of interpretation are often discussed in the interpreting community. These models propose to explain how interpreting is carried out mentally and model the underlying processes in the mind. Hence the term analysis of the interpreting *process,* in contrast to other analyses of the interpreting *product,* i.e., the text itself. Although this is only one level of analysis of interpretation process, it is frequently addressed in literature and there are several accepted models investigating this area in interpretation studies (Pöchhacker 2016: 79).

2.3.2. Information processing models

Models of cognitive processes underlying interpretation were put forward in the 1970s by David Gerver, and another by Barbara Moser-Mercer. Both are rooted in interdisciplinary research between psycholinguistics and interpreting studies (Moser-Mercer, 2002). Gerver’s model elaborated on the concept of a “short-term operational memory”, which is well-founded in psychological literature (Moser-Mercer, 2002). During simultaneous interpretation, as interpreters utter one part of the speech in the target language, they are listening to the next part of the source text and storing that information in their short-term memory for later interpretation. This model focuses on the flow and processing of information and simplifies the transfer from source to target language to “decoding-coding” of the source text. Thus, it does not touch on how background knowledge, context and long term memory helps understand, and transfer meaning (Moser-Mercer, 2002; Gumul, 2019). Moser-Mercer’s model incorporates the concept of long-term memory. Even though the interpreting process is largely helped by short-term or *working memory*, processes “occur in working memory with constant

access and feedback to long-term memory" (Moser-Mercer, 2002: 151). Her concept therefore stresses the importance of background knowledge stored in the long-term memory of interpreters, thus implies how preparation and contextual knowledge can also influence interpreting performance.

2.3.3. Gile's Effort Model

Daniel Gile (1995/2009) put forward a thoroughly researched model of cognitive processes, including aspects of memory. His aim was to understand how errors and omissions (*e/o*) happen during SI (Gile, 1995/2009), which he observed in numerous interpretations by professionals (Gile, 2017). The result is the Effort Model, describing four mental efforts during interpretation: Listening and analysis, Production, Memory, and Coordination.

Listening and Analysis is a conscious effort, because it involves distinguishing individual lexical units (words, numbers, names, etc.) from the perceived sound to analysing, understanding, and anticipating the meaning of the source text. (Gile 2009: 161) It is assumed that interpreters use more mental effort for listening and analysis than "regular" source text recipients. "Regular" source text recipients can direct their attention selectively to passages they find interesting, while interpreters must consciously listen and analyse throughout the entire source text. Moreover, interpreters' terminological and background knowledge is assumed to be less complete than that of source text recipients who are experts in the field. (ibid.: 162)

The Production effort "can be defined as the set of operations extending from the mental representation of the message to be delivered to speech planning and the performance of the speech plan, including self-monitoring and self-correction when necessary" (Gile 2009: 163). In other words, this effort involves interpreters' coping techniques in addition to mere speech production. What makes this a conscious cognitive effort is that interpreters are following speech structure and possibly, but not necessarily speech patterns of the speaker as opposed to how they would normally speak on their own. (ibid.: 163)

The Memory effort is based on literature about working memory, previously described in cognitive psychology and in Gerver's and Moser-Mercer's models of simultaneous interpretation. Although short-term memory is strongly interconnected

with long-term memory when the individual is carrying out tasks, “some authors even doubt the usefulness of the concepts of working memory as a separate entity” (Gile, 1995/2009: 167). Furthermore, there seems to be consensus that the memory effort is not automatic but requires mental processing capacity. (ibid.: 167)

It seems well supported with arguments that the above three efforts are (at least partially) non-automatic but require a conscious mental effort (Gile 1995/2009: 158-160), and that cognitive capacities at any given time have a limit. Therefore, there is a need for a fourth effort, the Coordination effort, to keep the other three within the available limit of mental efforts. Gile argues that faulty interpretations occur when the four efforts exceed the interpreter’s available cognitive capacity.

The Effort Model is directed mainly at explaining how common *problem triggers,* for example “names, numbers, enumerations, fast speeches, strong foreign or regional accents, poor speech logic, poor sound, etc.” (Gile 1995/2009: 171) cause mistakes in interpreting. Problem triggers may increase necessary cognitive efforts, thus leading to a cognitive overload even for professional interpreters and result in errors and omissions in the target text. Gile recognises this model’s limitation, that it is hard to link problem triggers to exact errors and omissions (e/o’s). For example, a sentence with poor speech logic increases the listening effort of the interpreter, which can result in limited mental capacities, thus e/o’s during the production phase. It is also possible that the interpreter has enough mental capacities to produce a complete target text from the sentence with poor speech logic. Nevertheless, the required production effort to do so can take away necessary cognitive capacities from adequately listening and analysing the next source text sentence. Thus, problem triggers may cause e/o’s even in the following text segments. Furthermore, Gile adds that they may not cause errors and omissions at all when coping strategies are applied. (ibid.: 172)

Therefore, Gile’s effort model has been criticised: it has not yet been clearly distinguished how cognitive overload leads to a quality decline in the SI product. In one of Gile’s own studies he tries to find a correlation between the effort model and interpreter’s performances:

> In a sample of 10 professionals interpreting the same source speech in the simultaneous mode, errors and omissions (e/o’s) were found to affect different source-speech segments,

> and a large proportion among them were only made by a small proportion of the subjects. In a repeat performance, there were some new e/o's in the second version when the same interpreters had interpreted the same segments correctly in the first version. (Gile, 2017: 153)

While it is possible that this is related to an inherent difficulty in simultaneous interpreting (the "tightrope hypothesis", Gile 1995/2009: 182-183), it is also possible that the observed mistakes were caused by human factor: the people making mistakes could have been mentally fatigued already before interpreting. They could have been less familiar with the topic, thus less prepared than the others. Thus, each interpreter's mental fatigue is an important aspect when evaluating cognitive loads during interpreting. Mental fatigue needs to be tested, their familiarity with the topic established, and their preparation methods considered. This can provide an additional information to the mere analysis of problem triggers, mental efforts and errors/omissions.

It is also worth mentioning that problem triggers have different degrees of difficulty. Some will maybe not cause as many difficulties as others. To illustrate this, Desmet et al. (2019) classify numbers according to difficulties in interpretation: *simple* numbers containing one or two meaningful units, e.g., five, sixty-seven, or six thousand, *linguistically complex* numbers containing more than two meaningful units, e.g., two hundred fifty million, *decimal numbers* e.g., 2,5, and *dates*. Therefore, having a problem trigger in the text will not necessarily cause problems in the interpretation.

2.3.4. Seeber's Cognitive Load Model

Seeber (2011, 2017) put forward another model about how different cognitive processes are carried out simultaneously during SI. He bases his model on detailed descriptions of linguistic and psychologic literature about working memory, language processing and comprehension. His Cognitive Load Model depicts similar ideas to Gile's Effort Model, in that listening and comprehension, production and monitoring, and information storage in working memory are conducted in real time and therefore interfere with each other, putting cognitive demand on simultaneous interpreters. (Seeber 2011: 187-189) When analysing potential problem points in interpretation, he mentions language specific difficulties, such as verb-final structures in German, as well as language independent difficulties such as a high speech delivery rate. The novelty of his approach lies in the underlying Multiple Resource Model of Wickens

(1984). “The Multiple Resource Model assumes that tasks interfere with each other more strongly when they have structures in common, i.e., if they demand the same level of a particular processing dimension” (Seeber 2011: 187). Processing dimensions are described as either verbal or spatial, and tasks are described as either perceptual, cognitive or response tasks. Since both *listening and comprehension* of the source text and *production and monitoring* of the target text in SI are verbal-perceptual and verbal-cognitive tasks[3], they interfere strongly with each other according to this model, hence the cognitive difficulty experienced in SI. (ibid.: 189)

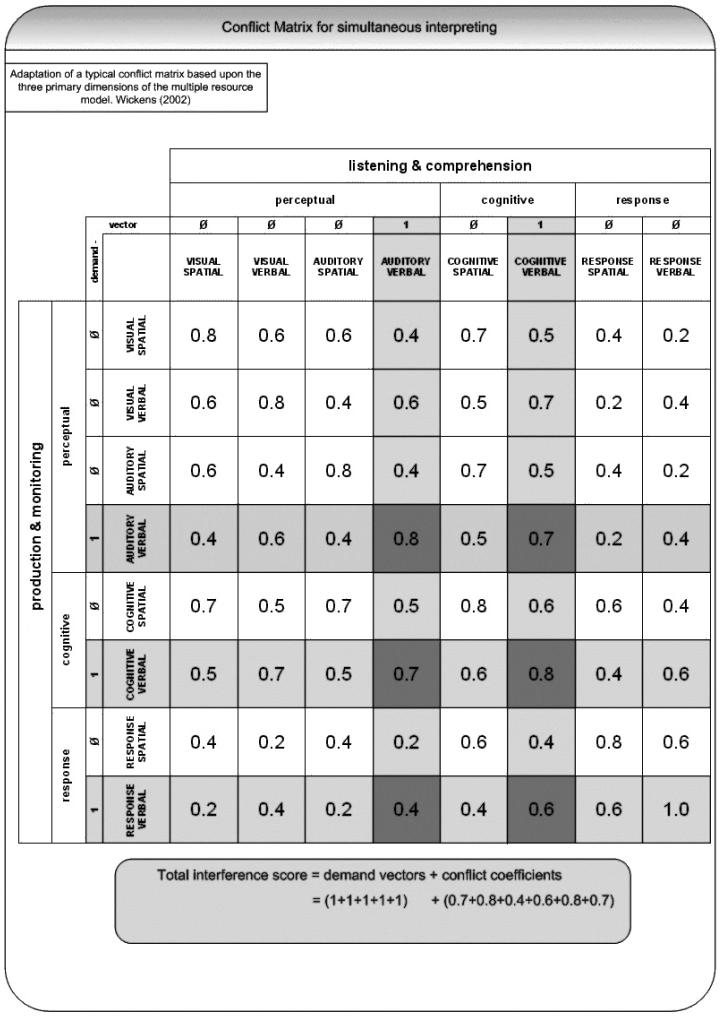
Seeber argues, contrary to Gile, that SI does not always exhaust all the available cognitive capacity of interpreters, since a) the SI tasks are not constantly overlapping, and b) experienced interpreters use different coping strategies to handle the interfering tasks thus freeing up cognitive capacities (ibid.: 190-196). This argument might well explain how interpreters have enough mental capacities to pay attention to their boothmate’s notes, follow a manuscript of the speech or even consult a glossary or word list during interpreting in real time. While it is hard for the psychologically untrained eye to comprehend how the values in the below Conflict Matrix are determined to be exact numbers, he is citing psychological studies upon which his findings are rooted. (Wickens, 1984, 2002)

However, Seeber does not entirely explain how the interference between different channels of attention seen in the figure below can be quantified. For example, what does a 0.8 conflict coefficient mean, when one is carrying out an auditory verbal *listening & comprehension* task and an auditory verbal *production & monitoring* task? The numbers suggest that these mental resources are conflicting each other: i.e., it becomes harder to carry out an auditory verbal *listening & comprehension* task and an auditory verbal *production & monitoring* at the same time, than it would be to carry out each one individually. This argument seems reasonable. Nevertheless, it is not mentioned whether the numbers adapted from Wickens’ research correlate to any experimental evidence.

It is also unclear how one should interpret the overall *conflict coefficient score*. Prandi (2019) adapts this Seeber’s table to describe these values and calculates that the *conflict coefficient* for SI using an Excel sheet as glossary is 16,8 and using a CAI

[3] plus a verbal response dimension when uttering the source text

software as glossary is 14,8. Results of her study indicate that the average precision of terminology in the interpreting products was higher using InterpretBank than using Excel. (Prandi 2019: 51) However, this was only the case in 4 out of 6 participants. 2 out of 6 participants achieved higher terminological accuracy with Excel than with InterpretBank. Therefore, these findings could be further verified with more participants.



Conflict Matrix for simultaneous interpreting

Adaptation of a typical conflict matrix based upon the three primary dimensions of the multiple resource model. Wickens (2002)

				listening & comprehension							
				perceptual				cognitive		response	
			vector	Ø	Ø	Ø	1	Ø	1	Ø	Ø
		demand ↓		VISUAL SPATIAL	VISUAL VERBAL	AUDITORY SPATIAL	AUDITORY VERBAL	COGNITIVE SPATIAL	COGNITIVE VERBAL	RESPONSE SPATIAL	RESPONSE VERBAL
production & monitoring	perceptual	Ø	VISUAL SPATIAL	0.8	0.6	0.6	0.4	0.7	0.5	0.4	0.2
		Ø	VISUAL VERBAL	0.6	0.8	0.4	0.6	0.5	0.7	0.2	0.4
		Ø	AUDITORY SPATIAL	0.6	0.4	0.8	0.4	0.7	0.5	0.4	0.2
		1	AUDITORY VERBAL	0.4	0.6	0.4	0.8	0.5	0.7	0.2	0.4
	cognitive	Ø	COGNITIVE SPATIAL	0.7	0.5	0.7	0.5	0.8	0.6	0.6	0.4
		1	COGNITIVE VERBAL	0.5	0.7	0.5	0.7	0.6	0.8	0.4	0.6
	response	Ø	RESPONSE SPATIAL	0.4	0.2	0.4	0.2	0.6	0.4	0.8	0.6
		1	RESPONSE VERBAL	0.2	0.4	0.2	0.4	0.4	0.6	0.6	1.0

Total interference score = demand vectors + conflict coefficients
= (1+1+1+1+1) + (0.7+0.8+0.4+0.6+0.8+0.7)

(Figure 1, from Seeber 2011: 188)

2.3.5. Effects of CAI tools with ASR on memory efforts

Since many CAI tools for SI, and in particular the CASSIS software, support the interpreter's lexical memory, there is a need to examine if and how they aid working memory during SI. However, consulting any sort of additional written material during SI will result in added cognitive load, as it introduces an additional visual-perceptual

and visual-cognitive task. Rodriguez and Schnell (2009) add terminological problems that can arise during the four efforts of interpretation. Sound distortions can hinder the listening and analysis phase. Moreover, recalling terms may prove to require a higher memory effort than available during the production phase. Therefore they argue that CAI tools have to make quick term retrieval during interpretation possible, and the "visual format of the terminology record must help the interpreter locate information quickly" (Rodríguez and Schnell, 2009). Moreover, they stress that smaller databases can be more useful during each individual conference and speed up term retrieval more.

An adaptation of the Cognitive Load Model to SI with CAI tools (Prandi, 2017,2019) suggests that using a bespoke software developed for interpreters could reduce cognitive load compared to using word processing software, or paper glossaries, when in need to look up terminology during interpreting. This finding is based on a quick search function of CAI software, that minimises the additional visual-spatial and manual-spatial task required to consult glossaries. However, this study did not yet evaluate the effect of integrated ASR functions, which could in theory further reduce the cognitive load when using terminological aids, due to automated showing of terms pairs from the glossary.

2.3.6. Effects of CAI tools with ASR on the listening effort

There has been extensive research on how following a speech manuscript during SI (also referred to as SI+T) can affect the interpretation process (Setton and Motta, 2007; Cammoun *et al.*, 2009). However, in SI+T, the text is delivered beforehand, and speakers can considerably differ from the manuscript due to for example restricted speaking time. The live transcription, as described in Chapter 2.1.7., follows the source speech in real time, therefore research results concerning SI+T do not necessarily apply here. There is a need to research how the live transcription will affect interpreting. Possible factors that can influence the usability of a live transcription are erroneous or slow transcription and a cognitive overload posed upon the interpreter, who listens to and reads the source text for the first time (unlike in SI+T where the manuscript is usually read or glossed over before interpreting). Therefore, consulting a live transcription can potentially make the Listening and

analysis Effort harder. The questions arising from this concept are: will the interpreter be able to divert cognitive resources to the transcription when encountering potential problem triggers in the speech? With what word error rate can the ASR system transcribe numbers and names, which are common problem triggers? If these problem triggers are transcribed and interpreted adequately, is it because the interpreter looked at the transcription for help? If these details are interpreted correctly, the following text segments must also be analysed, since theory suggests that failure sequences due to cognitive overload might occur later in the text, even if a problem trigger was interpreted correctly.

2.3.7. Effects of CAI tools with ASR on coping tactics and dealing with problem triggers

A potential problem trigger is interpretating numbers. Automatically showing numbers in context using ASR could also relieve interpreters of the cognitive load required to store them in their working memory. Thus, they would reduce the cognitive load required to recall terms from short term memory. Using ASR technology during interpretation of numbers has produced promising results in a first experiment. “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%.” (Desmet, Vandierendonck and Defrancq, 2019). Albeit in this experiment, the numbers showed without context on the interpreter’s screen. Moreover, the study relied on a simulation of ASR rather than actual ASR software, thus there was no possibility for said software to erroneously transcribe numbers. The experiment contained 5 numbers of each tested type: “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)” (ibid.: 20) All the numbers shown were controlled beforehand for correctness. Therefore, this phenomenon needs to be further investigated using real ASR technology with numbers in context. While problem triggers have been shown to not necessarily affect the target text (Gumul, 2019), the above analyses can help us understand how, under what conditions CAI tools can be used to the benefit of interpreters.

An aspect of the discussed cognitive models are the *coping tactics* used to overcome cognitively demanding text segments. Coping tactics are a conscious way to deal with problem triggers in the ST (Gile 1995/2009: 201). As the CASSIS software

may affect cognitive processes in SI, this paper will also investigate if and how it changes the coping strategies used during SI. Gile (1995/2009: 201-210) lists the following frequently used tactics:

Comprehension problems:

a) *Delaying the response*
b) Reconstructing a segment from context
c) Using the boothmate's help
d) Consulting resources in the booth

Preventive tactics:

a) Taking notes
b) Changing the ear-voice span
c) Segmentation
d) Changing the order of elements in an enumeration

Reformulation tactics:

a) *Delaying the response*
b) Using the boothmate's help
c) Consulting resources in the booth
d) Generalising
e) Paraphrasing
f) Explaining
g) Reproducing the sound of names
h) Naturalization
i) Transcoding
j) Omission
k) Informing the listeners of a problem
l) Referring the listener to another information source
m) Parallel reformulation
n) (Switching off the microphone – although Gile admits this tactic is dated*)*

Reading the live transcription in CASSIS could be a coping tactic in more cases: it could help with *comprehension problems*, that is if the ASR system has correctly transcribed something which the interpreter did not understand, it can be

sight-translated from the transcription. Interpreters may loosely follow the transcription as a *preventive tactic*, in anticipation of problem triggers such as numbers. Looking at the term base matches would be an example of using resources in the booth as a *reformulation tactic*. This raises the question how reading the term base matches will affect other reformulation tactics. That is, will interpreters include all the terminology shown by the software? Or will they still paraphrase, explain, generalise, naturalise, transcode, or omit terminology? If so, that could be a sign of using these coping strategies to overcome cognitive overload. Or will they find a way to incorporate the suggested terms, but get an unnatural sounding target text segment as a result? That could indicate how the additional visual-cognitive task of reading and comprehending the term on screen interfered with other cognitive tasks during SI. While quantifying cognitive loads is very challenging, some inferences can be made by analysing SI transcripts (Setton, 2011) and triangulating the found errors and coping strategies with retrospective interviews, and other methods measuring stress and mental fatigue of study participants (Mellinger 2019: 37).

3. Research question and thesis outline

A survey by Corpas Pastor and Fern (2014) identified a need to investigate computer tools made for interpreters. Specifically, the study authors recommend some sort of immediacy while using terminology tools, for example in automatised look-up functions of the programs *LookUp*, *Interplex UE* and *InterpretBank* (Corpas Pastor and Fern, 2016: 35). Fantinuoli (2018) has put forward a way to further accelerate the process of looking up terms using automatic speech recognition (ASR). This development could be a possible response to the inadequacy of interpreting software solutions perceived in the study of Corpas Pastor and Fern as they were recommending some degree of automation. However, the development of computer-assisted interpreting could impose other effects on the workflow of interpreters as seen on the example of the changed workflow of translators after the arrival of CAT tools and other software (Pym, 2011). Stewart et al. (2018) propose that interpreting software with ASR could increase the cognitive loads during interpreting by presenting too many term base matches too quickly. One can therefore argue that the changes towards automation are not necessarily good or bad, but rather create new kind of challenges which need to be understood and investigated in detail.

This thesis will focus on software for interpreters that utilise ASR and investigate the CASSIS software in detail. One of the main features of the software is to automatically present interpreters with suggested target language terms based on a bilingual list of terms uploaded before interpretation.

The first research question regarding SI product is: How does this software influence terminological accuracy? Methodology will be based on Prandi's exploratory study design (2019). Participants will interpret a 10-minute-long English speech into German containing specialised terminology. 5 minutes will be interpreted with CASSIS, 5 minutes without it. Terminological accuracy will be determined according to the classification in Prandi's study design.

CAI tools with ASR that transcribes numbers has been shown to improve the accuracy of numbers in interpretation as well (Desmet, Vandierendonck and Defranq, 2019; Defrancq and Fantinuoli, 2020). Thus, the number accuracy during interpretation will also be discussed. The types of errors and inaccurate renditions will be evaluated and the effects of the CAI tool on them will be discussed.

The preparation phase for an interpreting assignment and interpreter's background knowledge on the topics of the source material can affect the interpreting product (Moser-Mercer, 1992; Will, 2007; Rodríguez and Schnell, 2009; Xu, 2018). Participants will thus be asked to reflect on their terminological preparation prior to the study and how CASSIS term base affected their terminological performance. Other external effects such as previous interpreting experience, working language combination of interpreters will also be assessed and considered when discussing the results.

Lastly, the paper will investigate the implications of CASSIS on cognitive processes in the booth, particularly on the coping strategies used. (Strolz, 1994; Gile, 1995/2009; Seeber, 2011) The evaluated terminological accuracy of renditions will be cross-referenced with a fatigue test of participants and a follow-up questionnaire (Ivanova, 2000) to gauge cognitive loads, problem triggers and adaptation of coping strategies when using CASSIS.

Evaluating the findings is expected to shed light on whether CASSIS changes the workflow of translators, the terminological accuracy, and the interpreting process.

4. Methodology

4.1. Participants

There were 6 participants in this study in total. 5 participants of this study were advanced MA interpreting students from the University of Vienna and one of them was an interpreting student of ITAT Graz. They have all successfully completed at least entry level interpreting courses and thus already have practiced simultaneous interpreting. English is their A, B or C language and German either their A or B language.

4.2. Recording methods

The study simulated simultaneous interpretation with the CASSIS software and was set up remotely due to a series of restrictions imposed by the COVID-19 pandemic. It was set up as follows: The source speech video was played back on the source device, for the sake of simplicity *device A*[4]. Device A provided source speech audio input to *device B*[5] via a 3,5mm AUX cable and an external soundcard. This means very high audio input quality for running CASSIS on device B. Device B runs CASSIS which participants can see via Zoom. Participants enter the remote meeting one by one. They saw the screen of device B and listened to the speech simultaneously. Participants then interpreted the speech and recorded their renditions separately. The CASSIS software screen was also recorded using the OBS software, to triangulate participants renditions with the software's visual interface.

4.3. The source speech

The English source speech is an outtake of the second State of the Union Address 2021 delivered in the European Parliament by Ursula von der Leyen. The full transcript of her speech is available at https://ec.europa.eu/commission/presscorner/detail/ov/_SPEECH_21_4701 and the German translation at https://ec.europa.eu/commission/

[4] Device A would correspond to the booth equipment in a real-life conference interpreting situation, which provides interpreters the source speech audio.

[5] Device B would correspond to the interpreter's laptop, or a PC in the booth provided by the conference organizer. An internet connection on this device is a pre-requirement for running CASSIS.

presscorner/detail/de/SPEECH_21_4701. Participants were not aware of the speech in question before interpretation, nor about the existing transcripts. It is possible, however, that they were nonetheless familiar with the speech beforehand. This was also surveyed in the questionnaire and considered when evaluating the data.

Participants interpreted a segment of the above speech entitled A EUROPE UNITED THROUGH ADVERSITY AND RECOVERY. The segment is available as a video at https://webgate.ec.europa.eu/sr/node/31080. There are no slides or other types of visual information in this video for interpreters to rely on, apart from the body language of the speaker. The topic until minute 5:00 is the health crisis caused by the Covid-19 pandemic, and from minute 5:00 till minute 11:30 the resulting economic crisis and production shortages.

There is a period of rhetorical break and applause from 5:00 till 5:05 which coincides with the shift in speech topic. Recording and interpreting were also briefly paused at this point during the study. This break was used to change over from interpreting with a paper glossary to CASSIS with half the participants, and vice-versa with the other half. The speech was then interpreted from minute 5:05 until minute 10:05. This counterbalancing order provides data using the equal lengths of speech in each format. It also ensures that mental fatigue towards the end of the interpretation task is to some extent equalised between the group testing CASSIS and the control group. The video from minutes 0:00 till 10:05 was transcribed in the appendix of this paper.

The speech is paced at 120 words per minute (WPM) with negligible tempo fluctuations: minutes 0:00 till 5:00 have a tempo of 123 WPM, while minutes 5:05 till 10:05 have a tempo of 117 WPM. Since 120 WPM is regarded as the upper threshold of comfortable pace for SI (Riccardi, 2015), speech rate alone is not expected to act as a problem trigger. Moreover, it is delivered with an easy-to-understand, albeit non-native accent. The convenient speech tempo and accent are planned to ease the Listening and Analysis effort enough to free up mental capacities for searching the glossary or looking at CASSIS.

Accuracy of the underlying ASR engine could be affected by the non-native accent of the speaker. Its accuracy can be evaluated according to the following formula: Word Error Rate (WER) = total number of substitutions, insertions, and

deletions in transcription divided by total number of words in the original sequence. The total word count until minute 10:05 is 1203 words, whereas the total error count was 89. This resulted in a WER of 7,4%, while error rates of 5% or below are considered a good value for ASR systems (Chapter 2.1.8). These mistakes could cause misunderstandings when someone is following the live transcription during SI. The first half of speech until 05:00 was transcribed slightly better with a WER of 7,3% which means 45 errors out of 613 words. In the second half (from 05:05 until 10:05) the ASR engine achieved a WER of 7,4% with 44 errors in 597 words. In terms of numerals however, the ASR system reached an error count of 0, as it transcribed all numbers from the speech accurately.

4.4. Relevant specialised terminology

The speech as delivered contains terms from the following subject fields: politics, healthcare, economics, and technology. The terms below correspond to the main subject fields of the speech. Moreover, they have all been translated into German in the official press release. “This indicates that in a time-unconstrained scenario, the term *should* be translated.” (Vogler, Stewart and Neubig, 2019: 110). Of course, one cannot expect all these terms interpreted during SI as some of these terms will likely be interpreted using coping strategies - e.g., generalisations, paraphrasing. Also, some of these terms will possibly be omitted in SI products even by the most experienced interpreters due to problem triggers. However, predicting how and to what extent these coping strategies would have be applied during SI was impossible since it is interpreter-dependent. Therefore, the written translation found online was the basis of relevance for this list of terms and thus for the CASSIS bilingual term list.

A list of these terms was sent to participants two days before translation. They were informed about the topics one week ahead, on which they could base their own glossaries. However, the bilingual term list used by the CASSIS software during the study was only based on the below word list:

List of terms			
	minutes 0:00-5:00	minutes 5:05-10:05	in total
politics	14	12	26
	geopolitical issues, Team Europe, Commission, member states, our neighbourhood, European Health Union, proposal, HERA authority, competent national authorities, funding, to propose a mission, house (of EU parliament), EU Digital Certificate	SURE, structural dialogue, at the top level, Economic Governance Review, build consensus, ambitious proposal, gatekeeper power, democratic responsibility, make-or-break issue, digital spending, Next Generation EU, European tech sovereignty	
healthcare	15	0	15
	pandemics, vaccine, Covid-19, fully vaccinated, dose, to administer (a dose of vaccine), mRNA, global health, vaccination rate, booster shot, pandemic of the unvaccinated, pandemic preparedness, health threat, health preparedness and resilience, epidemic		
economics	5	25	30
	low-income countries, production capacity, scientific capacity, private sector knowledge, return on investment	Eurozone GDP, pre-crisis level, pre-pandemic level, Euro area, to outpace (competitors), quarter, financial crisis, short-term recovery, long-term prosperity, labour market	

		reforms, pension reforms, tax reforms, increased debt, sectors, new ways of working, single market, (market) erosion, (market) fragmentation, digital single market, to foster innovation, production lines, demand, shortage, value chain, manufacturing capacities	
technology	2	6	8
	digital certificate, plugged in	5G, fibre, digital skills, artificial intelligence, semiconductors, smart factories	
total terms	36	43	79

Enumerations were split into their component terms and renditions of their component terms were evaluated during the study. Furthermore enumerations, as potential problem triggers, were asked about during the retrospective interviews. The speech contains the following enumerations:

Enumerations			
	minutes 0:00-5:00	minutes 5:05-10:05	in total
	2	6	9
containing:	vaccinations, vaccination rates, pandemic preparedness// scientific capacity, private sector knowledge, competent national authorities	short-term recovery, long-term prosperity// labour market reforms, tax reforms, pension reforms// increased debt, different sectors, new ways of working// gatekeeper power, democratic responsibilities, to foster innovation, artificial intelligence// smartphones, electric scooters, trains,	

		smart factories// value chain, design, manufacturing capacities	
	6 component terms	19 component terms	25

4.5. Product analysis

4.5.1. Terminological accuracy

Transfer of the listed terms into German was measured according to the following three categories intended for measuring simultaneous interpretation with CAI software:

> *"**Close renditions** (precision 2 – P2): no information lost, precise rendition, use of equivalent as per glossary or adequate synonym*
> ***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)."* (Prandi, 2019: 50)

Adding up the terminological accuracy scores from the first 5 and from the last 5 minutes shows the terminological accuracy score for the paper glossary and the CASSIS system respectively. Comparing the precision scores will provide data about the way CASSIS affects terminological accuracy compared to using a paper glossary.

To provide another level of analysis, *acceptable, and zero/unacceptable* renditions were assessed to see what kind of *coping tactics* (Chapter 2.3.3.) were applied when encountering these difficulties. Analysing these could provide another layer of data about how CASSIS affects the interpreting product. The reasons behind using coping tactics were examined during the retrospective interviews.

4.5.2. Accuracy of numbers

Firstly, accuracy score of numbers is calculated using the system described above of *accurate* (P2), *acceptable* (P1), and *zero/unacceptable* (P0) renditions. Accuracy is compared between the two different kinds of glossaries used in this study.

Score 2 is given only to accurate renditions since synonyms are not applicable here. Rounding up or down and approximation within the same order of magnitude are counted as *acceptable* (P1). For example, acceptable renditions for 47 would be “more than 40”, “less than 50”, “approximately 40” or “approximately 50”. Generalisations within reason are also counted as *acceptable* renditions, e.g., in case of 47 countries “numerous countries” would be *acceptable*, while “a few countries” would be counted as *unacceptable* (P0). Desmet et.al. (2019: 21) enlist further possible mistakes resulting in *zero/unacceptable* renditions: *omission, transposition* (e.g., 47 becomes 74), which is a possibility, since their lexical units are pronounced in opposing order in German, *wrong order of magnitude* (e.g., million becomes billion), *phonological mistakes* (e.g., 14 becomes 40).

Desmet et. al. (2019) categorise numbers according to their difficulty during SI: *simple numbers* containing one or two lexical units are the easiest to interpret. They define *complex numbers* as having more than two lexical units and therefore more units for interpreters to recall or note down. *Decimal numbers* and *years* are also more difficult to interpret simultaneously than simple numbers. The speech used in this study contains more simple numbers than the other three categories combined.

The accuracy gain using ASR of simple numbers was smaller than the accuracy gain with the other three number types in the study conducted by Desmet et.al. During the present study, all numbers were considered potential problem triggers, their renditions were therefore all analysed. The speech contains the following amount of *simple, linguistically complex, decimal numbers,* and *dates*:

Numbers			
	minutes 0:00-5:00	minutes 5:05-10:05	in total
simple	10	4	14
complex	5	1	6
decimal	1	1	2
year	1	1	2
total	17	7	24

4.6. Process analysis

4.6.1. Questionnaire

Before the interpreting task, interpreters filled out a questionnaire about their use of CAI tools. This shows how familiar they are with the mentioned tools, and whether they have tried something similar before. If an interpreter is already familiar with CAI software, that could affect their performance while using the software. On the other hand, if they are not familiar with them at all, or if they have a sceptical view of said software, that could also affect the results, therefore this aspect must be considered.

These questions, asked before interpreting, are adapted from a usability survey for CASSIS made by the software developer. The questions were adapted to fit the purpose of this study and presented to the study participants in English (see Appendix) and assess the below aspects.

-Basic questions to assess interpreter demographics: age, gender, interpreting experience, language combination.

-Previous experience on CAI tools: Tools and/or software used in previous assignments both during preparation and during SI. Tools and/or software used in preparation for this study.

The results of this questionnaire are considered during evaluation of the interviews and performance.

4.6.2. Retrospective interviews

Interpreters were asked about their experience while encountering the limits of their cognitive capacity, high mental loads, and problem triggers during SI (Chapter 2.3.3. Gile's Effort Model and Chapter 2.3.4. Seeber's Cognitive Load Model). The interviews were conducted immediately after the SI task was completed, given that close temporal proximity to the assignment aids retrospective evaluation of participants (Ivanova, 2000). The full transcript of the source speech was shown, as "presentation of the original stimuli in conjunction with use of the observational notes" (ibid.: 33) can effectively prompt retrospective memory.

The retrospective interviews were used to evaluate perceived usability of the tested tool during high cognitive loads, as well as the perceived effect of cognitive overloads on problem triggers and coping strategies. The questions therefore are:

-Did the interpreter look at CASSIS, or the paper glossary, when encountering problem triggers, especially numbers and enumerations? If yes, what kind of influence did it have on the interpreter's cognitive capacities?

-Did the interpreter look at CASSIS, or the paper glossary, for terminological help, and with what results? Did the software display the terms they were looking for? What kind of influence did it have on the interpreter's cognitive capacities?

-Did mistakes in the ASR transcription affect usability, or cause distractions?

4.6.3. Surveying interpreter preparation

One of the key areas of this study is to investigate the effects of the CASSIS software on terminology. Preparing for an assignment has an impact on how accurately interpreters are performing (Díaz-Galaz, Padilla and Bajo, 2015). Interpreter preparation was therefore considered while evaluating terminological performance, and the use of coping tactics.

Keeping in mind the crucial role of preparing for an assignment in the three-stage model (Will: 2007), the terminological preparation of participants was assessed. Stage I of preparation is aimed at developing a general understanding, and background knowledge in the subject field(s), which supports understanding of the source text. All 6 interpreters were given one week to prepare their background knowledge in the subject fields of the speech. They were sent the list of specialised terminology contained within the speech two days ahead, to simulate the reception of relevant conference documents during Stage I of the terminological preparation model by Will. The word list gave interpreters a more specific idea about the direction of the source speech, than just the main speech topics. This opened some possibility for different levels of preparation between interpreters, which was accounted for within the pre-study questionnaire.

Before interpreting, participants were asked to evaluate on a scale of 1-10 where 1 is lowest and 10 is highest, how well they understand the special domains contained in the speech: EU politics, medicine/health, economy, and technology, and

how well they think they have prepared for this assignment. This evaluation was used as an additional level of data supporting the interviews after SI.

Well-prepared interpreters with a sound background knowledge of the speech topics could potentially use CASSIS differently than those who are not confident in their preparation.

4.6.4. Mental fatigue test

A mental fatigue test was conducted to see how the CASSIS software affects cognitive capacities of users. Using the live ASR transcription when encountering problem triggers could affect the Listening and Analysis effort, as it presents an additional visual input to the aural input.

Firstly, mental fatigue was self-assessed before interpreting, at the 5-minute break and at the end on a scale of 1-10. Verschueren et al. (2020) have studied mental fatigue and found that participants performing a challenging task have assessed their mental fatigue lower than the control group.

Participants have also performed a verbal memory test. This test is aimed at quantifying their mental fatigue in terms of verbal memory. It is a test available online and shown to participants only during the experiment, available at https://humanbenchmark.com/tests/verbal-memory. The memory test consists of individual English words shown one by one on their screen. Some of the words will repeat, but not all of them. Participants will then have to answer whether they have seen the word before or not. Participants have three possible wrong answers and the third one concludes the test. Even though this mental fatigue test aims at being more objective than the self-assessment of each participant, it is still not totally objective. The words and the repeating words are randomised, which could result in difficulty differences between each try.

4.7. Analysis of recordings

The changes in coping tactics according to Gile (1995/2002) were also investigated. Each participant gave an interview about how they used their resources during the experiment: how often they looked at the glossary, how often they looked at the software, whether they noted numbers, how much they relied on their preparation and interpreting skills and with what result they used these resources.

Each interpretation was then time-matched and edited into one video with the screen recording of CASSIS. These recordings served as a basis for cross-examination with the interviews. The point of cross-examination was to determine which coping tactics the interpreters used sentence by sentence, if there were any coping tactics identifiable.

The following coping tactics were counted from the recordings: delaying the response, reconstructing a segment from context, changing the ear-voice span, segmentation, changing the order of elements in an enumeration, generalising, paraphrasing, explaining, naturalization, transcoding, omission. This analysis tries to determine how using the CASSIS tool has had an impact on the interpretation product. Due to the low number of participants the findings may be unrepresentative.

The following is an example of how recorded interpretations were analysed. Two sentences from the source speech at 00:42 are: "*We delivered more than 700 million doses of vaccines to the Europeans. We delivered, indeed we delivered also more than 700 million to the rest of the world, to more than 130 countries."* The following coping tactics adapted from Gile are marked in the quoted German interpretation, highlighted with the following lines: changing the ear-voice span, delaying the response, generalization, omission. Pauses of about 1 second are marked with 2 slashes: „**". Here is the interpretation of study participant no.4: „*Wir haben über 700 Millionen Impfdosen ** wir haben trotz allem ** fast 700 Millionen in viele Staaten der Welt geliefert."* The interpreter omitted these 3 words: "*to the Europeans"*. The interpreter then chose a delaying tactic during the repetition "*we delivered, indeed we delivered*", possibly to gain more details before inserting the infinite German verb for "*delivered*" which, according to German grammar, must be the last word of the sentence in the German past tense used here. The interpreter has delayed her response to listen to the object and further details in the source sentence before they interpret the verb into German. She also changes the ear-voice span by approximately one more second. As the destinations are uttered: "*to the rest of the world, to more than 130 countries"*, the interpreter generalises that as "*in viele Staaten der Welt*" and then inserts the infinite German verb for "*delivered*" to finish the sentence. There was one omission, one delay of response, one change of ear-voice span, and one generalisation each.

The coping tactics were not counted per word, but rather per continuous unit of text where a coping tactic was used. For example, in the previously quoted segment the words "*to the Europeans"* were omitted from the German interpretation. Since three continuous words were omitted, this was counted as the tactic omission used 1 time. The following sentence is at 04:22 in the source text: "*The work on the European Health Union is a big step forward and I want to thank this house for your support*." Interpreter no. 4 said: "*Die Arbeit der europäischen Gesundheitsbehörde ist ein sehr großer Schritt der Union.*" Since the expression of gratitude towards members of parliament was not interpreted, it counted as 1 omission of meaningful part of text, even though the source sentence segment "*and I want to thank this house for your support"* was made up by 9 words. In other words, the number of times a certain coping tactics was used, was based on continuous parts of text carrying meaning or speaker intention. Since this categorisation is rather subjective, the analysis of coping tactics has a qualitative rather than quantitative character.

How often the interpreters were consulting resources in the booth and taking notes was only estimated based on the time-matched recordings and the interviews after interpretation. However, no exact number can be determined from this experiment, since no eye-tracking device was available. Participants' use of paper-based resources could not be recorded either due to the remote experiment, only estimated based on the interviews. A recording with a web camera would have jeopardized their anonymity and would not have delivered enough input to see whether they have noted down something, looked at their notes, or looked somewhere unrelated. There was not enough time within the experiment to ask participants about every single coping tactic and every single issue they encountered. Therefore, the below data is a qualitative rather than quantitative analysis and has potential to be improved in further experiments.

If for example, the interpreter was heard pausing during an enumeration, then incorporating the matches from the term base into their interpretation after they were shown up in the right-hand column of CASSIS, the following data points were added: 1 for changing the ear-voice span, 1 for consulting the resources in the booth and further data points under the other coping tactics if applicable. The following example is an interpretation of a sentence at 05:35 in the source text: "*Last time it took us eight years for the Eurozone GDP to get back to pre-crisis level. This time, we expect 19*

countries to be at pre-pandemic levels this year (…)." Interpreter no. 6 audibly changed the ear-voice span at the underlined spot and then incorporated a match marked with a double underline that was shown on the right-side column of CASSIS at the same time: „*Es hat 8 Euro/ 8 Jahre * gedauert, um das BIP in der Eurozone auf das Vorkrisenniveau zu heben und diesmal waren es 19 Länder, die * schon ** öhm, nach einem Jahr, das Niveau vor der Pandemie erreicht haben*." The terms marked with a dotted line were also shown up in CASSIS, however there was no hesitation sound before, which is why it was estimated that the interpreter has recollected these from memory. Of course, this is an assumption based on the analysis of recording and the interview where this interpreter reported to having used the suggested translations from CASSIS at times when she could not recollect an appropriate term from memory. It is possible that the interpreter was not looking at CASSIS but only delayed their response to have more cognitive capacities available to recall the term from memory. Therefore, the data about how often the interpreters consulted resources in the booth as a coping tactic, is subject to some assumptions and could be potentially incomplete.

5. Quantitative study data

5.1. Participant demographics

6 participants have completed the study. Their demographic information is below:

1. Gender

   3 male, 3 female

2. Age

   24, 27, 27, 28, 28, 30

3. Working languages:

   German: 1x A language, 5x B language

   English: 1x A language, 1x B language, 4x C language

4. Studies

   5 have studied both consecutive and simultaneous interpreting on MA level

   1 MA student of consecutive interpreting with a completed entry level simultaneous course

5. Interpreting experience

   4 advanced MA students, 2 MA graduates with less than 5 years of professional experience

5.2. Terminology accuracy scores

In the process of evaluating the terminology accuracy scores of the 79 selected terms based on the scoring system of Prandi (2019), participants received 2 points for *close renditions*, 1 point for *acceptable renditions*, and 0 for *zero/unacceptable renditions.*

The sentence "*We have 1.8 billion additional doses secured."* – contains the medical term *doses* [of vaccine]. Examples of close renditions for this term included the German terms: *Impfdosen, Dosen, Impfungen*. These were counted as 2 points. The German term *Impstoff* was counted as an acceptable rendition for 1 point as it is a change in unit: instead of one dose of vaccine it means the type of vaccine in general. Omissions of the term were counted as 0 points.

Compound terms such as "*mRNA production capacity"* were given a maximum of 2 accuracy points for the whole. Examples of close renditions for 2 points include: "*mRNA-Produktionskapazität*", "*mRNA-Herstellungskapazität*". An example of an acceptable rendition for 1 point is: "*Fertigungskapazität*" which stands for "*production capacity"*. Even though the medical term *mRNA* was missing, the previously interpreted sentences provided enough context to understand that the production capacities have to do with vaccine production.

The terminology accuracy scores are below:

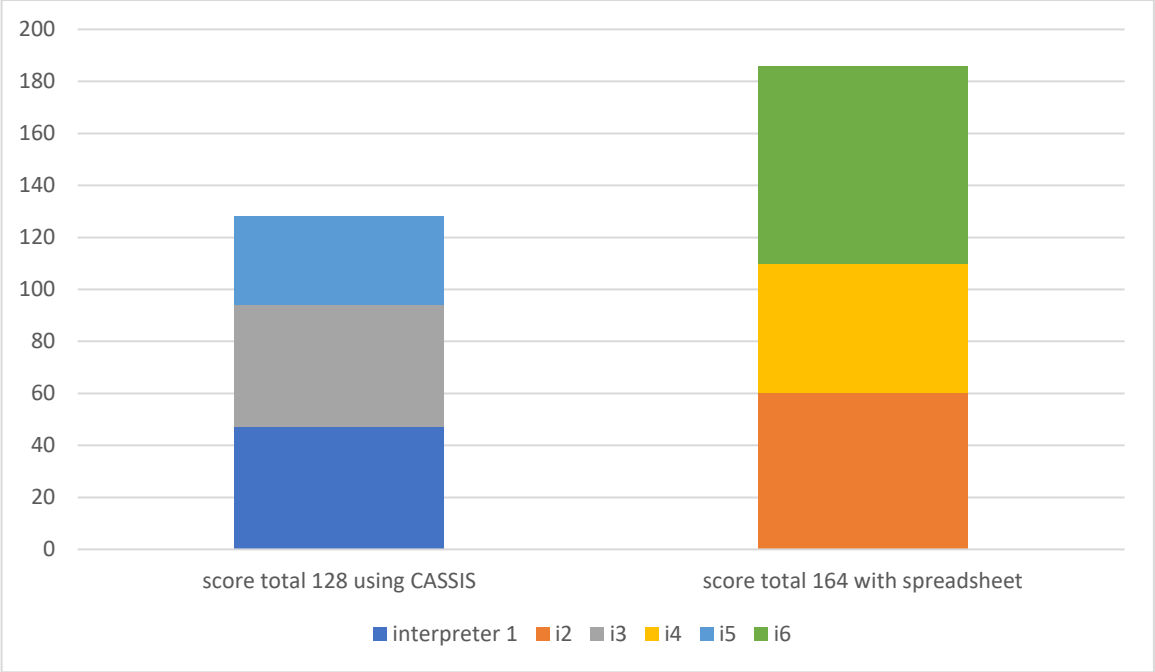
Participant	**1**	2	**3**	4	**5**	6	average
CASSIS used at min. 0-5	**Yes**	No	**Yes**	No	**Yes**	No	
Terminology accuracy score	**47**	56	**47**	48	**34**	60	48,67

Table 1. terminology accuracy scores at minutes 0-5

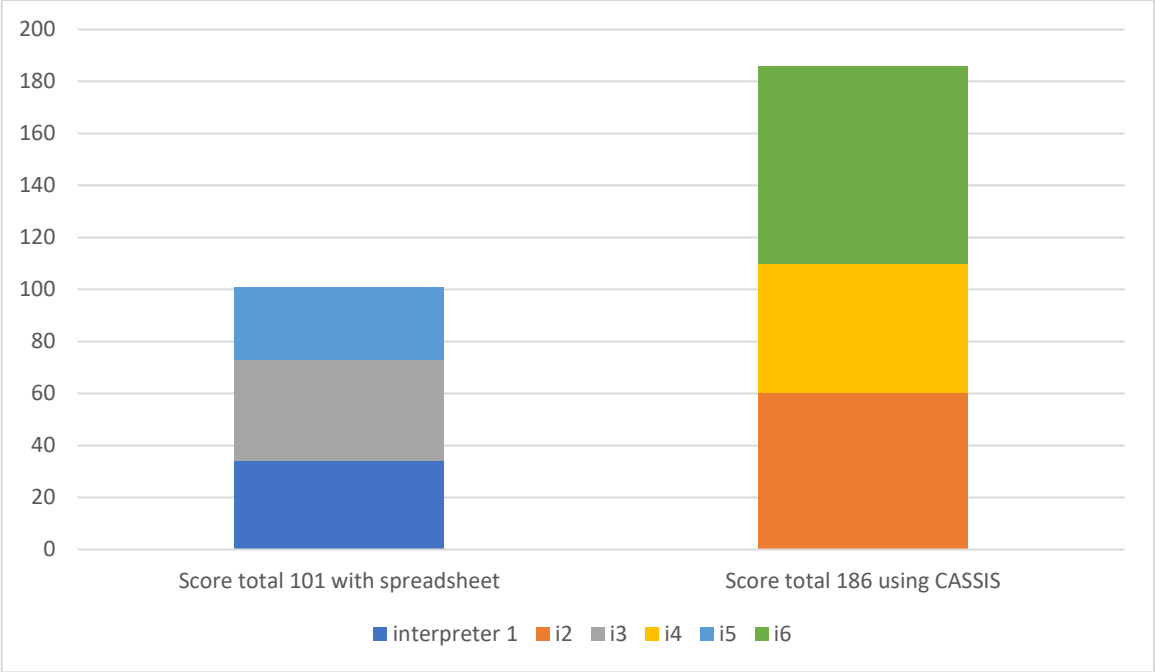
Participant	1	2	3	4	5	6	average
CASSIS used at min. 5-10	No	**Yes**	No	**Yes**	No	**Yes**	
Terminology accuracy score	34	**60**	39	**50**	28	**76**	47,83

Table 2. terminology accuracy scores from minutes 5-10

The values from Table 1. and 2. are visualised per each glossary method below. The control group has used a spreadsheet format glossary either in Excel or printed out.



Graph 1. terminology accuracy scores at minutes 0-5 per glossary method



Graph 2. Terminology accuracy scores 5-10 per glossary method

The overall scores are as below:

Interpreter	1	2	3	4	5	6	average
score	81	116	86	98	62	136	96,5

Table 3. Terminology accuracy scores 0:00 – 10:05

5.3. Terminology score differences between the two glossary methods

The maximum achievable terminology accuracy score is 72 for the first half of speech and 86 for the second half. The score percentages for both halves of speech can be calculated from these maximum values. The below tables include how many percents of the maximum accuracy score each participant achieved.

Interpreter	**1**	2	**3**	4	**5**	6	average
CASSIS used	**Yes**	No	**Yes**	No	**Yes**	No	-
Percentage at min. 0-5	**53,41%**	63,64%	**53,41%**	54,55%	**38,64%**	68,18%	55,3%

Table 4. terminology accuracy percentages during minutes 0-5

This means an average terminology accuracy percentage of 48,48% with CASSIS and 62,12% with a spreadsheet glossary in the **first 5 minutes of the speech**, amounting to an average **difference of 13,64%** of accuracy of the selected terms **in favour of the spreadsheet** glossary.

Interpreter	1	**2**	3	**4**	5	**6**	average
CASSIS used	No	**Yes**	No	**Yes**	No	**Yes**	**-**
Percentage at min 5-10	33,33%	**58,82%**	38,24%	**49,02%**	27,45%	**74,51%**	46,9%

Table 5. terminology accuracy percentages during minutes 5-10

This means an average terminology accuracy percentage of 60,78% with CASSIS and 33% in the control group **from minutes 5 to 10 in the speech**,

amounting to a **difference of 27,78%** of accuracy of the selected terms **in favour of the CASSIS automatic glossary**.

Judging by the average values, the first half of text seemed to be somewhat easier as in average interpreters achieved 55,3% accuracy. They achieved on average 46,9% in the second half, which is on average 8,4% lower than the first half.

From tables 4 and 5, the accuracy percentage Δ [7] of each participant can be calculated between the two glossary methods. This calculation includes the whole interpretation.

This means an average Δ terminology accuracy of +7,07% when using CASSIS.

Interpreter	1	2	3	4	5	6	average
Terminology accuracy percentage with CASSIS	Higher than using a spreadsheet	Lower	Higher	Lower	Higher	Higher	Higher
Δ terminological accuracy	20,07%	-4,81%	15,17%	-5,53%	11,19%	6,32%	7,07%

Table 6. accuracy percentage change for each interpreter

5.4. Number accuracy scores

The number accuracy scores were evaluated based on the scoring system of Prandi (2019) and Desmet et al. (2019). Participants received 2 points for *close renditions*, 1 point for *acceptable renditions*, and 0 for *zero/unacceptable renditions.* The achievable accuracy points were the same regardless of complexity of numbers. Simple numbers such as "*30"*, and complex decimal numbers such as "*1.8 billion"* were worth 2 accuracy points each if interpreted correctly.

Examples of acceptable renditions of the number "*1.8 billion euro"* for 1 point each included approximations such as "*mehr als 1 Milliarde Euro*" which stands for

[7] Δ, or „delta" = the mathematical symbol for a difference between two values

more than 1 billion euro and generalisations such as "*eine erhebliche Summe*" which stands for *a substantial amount*. These renditions are not fully accurate but also not wrong.

Wrong orders of magnitude which resulted from the difference between German and English were however evaluated as unacceptable renditions for 0 points. Unacceptable renditions of "*1.8 billion Euro*" were in German: "*1,8 Millionen Euro*", or "*1,8 Billionen Euro*", since both are wrong by a factor of x1000. Similarly, swapping the digits due to the differences between German and English, e.g., "*forty-two"* being interpreted as "*vierundzwanzig"* which stands for *twenty-four*, also resulted in 0 accuracy points.

The scores were evaluated for both parts of text.

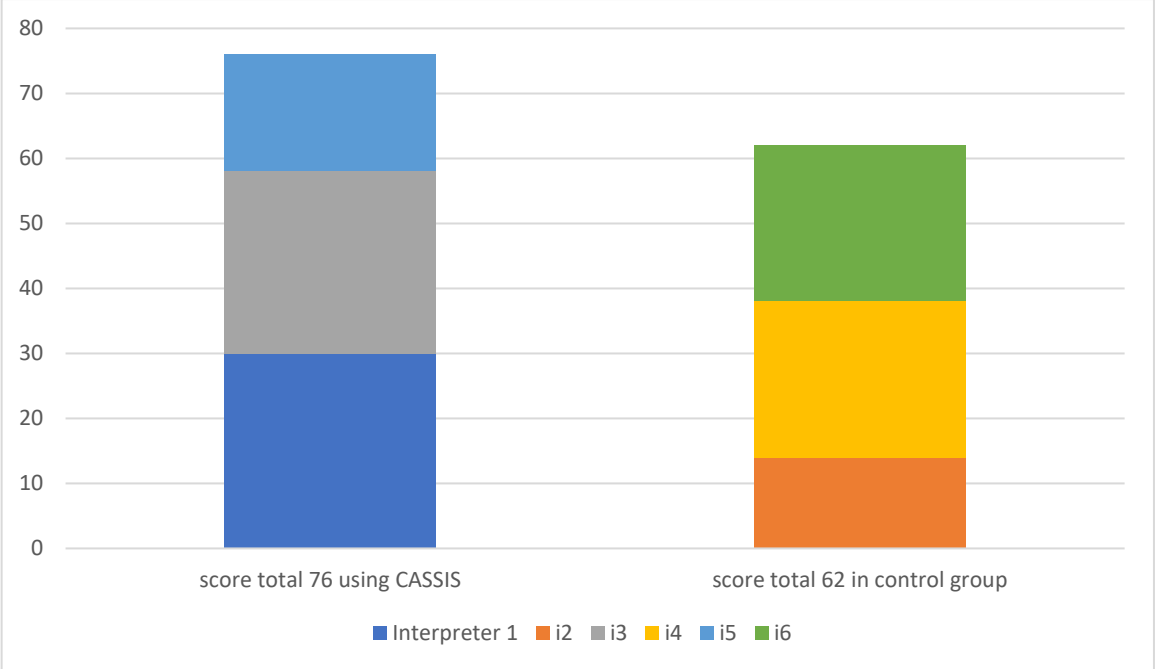
Interpreter	**1**	2	**3**	4	5	6	average
CASSIS used	**Yes**	No	**Yes**	No	**Yes**	No	
number accuracy score 0-5 min	**30**	14	**28**	24	**18**	24	23

Table 7. number accuracy scores during minutes 0-5

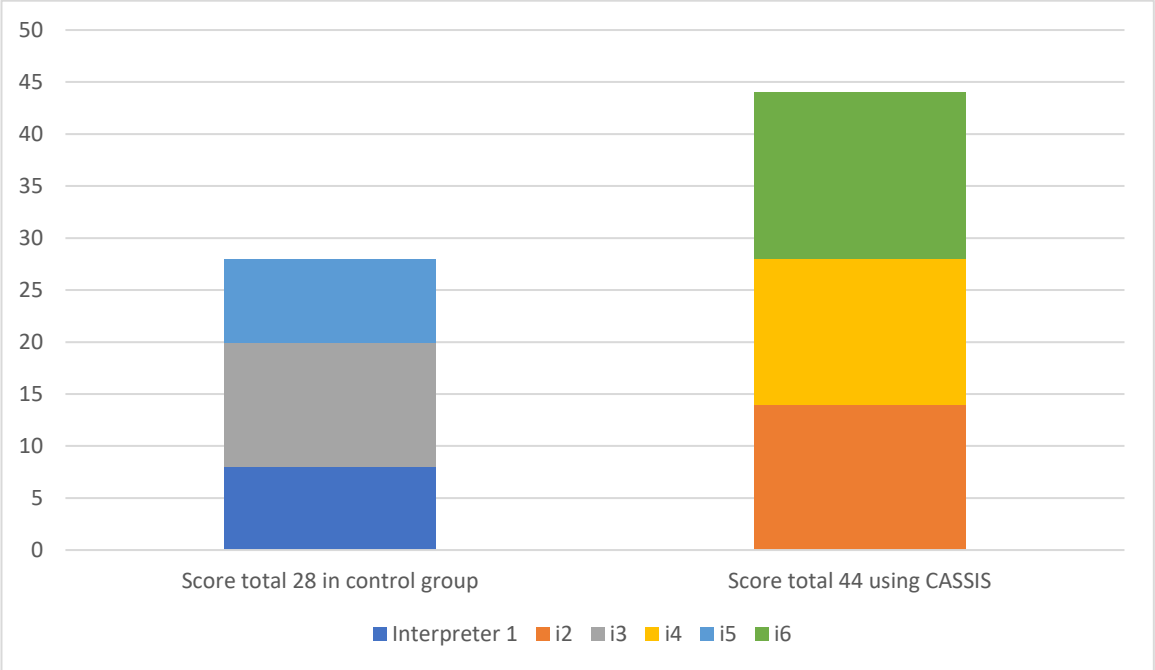
Interpreter	1	2	3	4	5	6	average
CASSIS used	No	**Yes**	No	**Yes**	No	**Yes**	**-**
number accuracy score 5-10 min	8	**14**	12	**14**	8	**16**	12

Table 8. number accuracy scores during minutes 5-10

The scores are visualised below:



Graph 3. number accuracy scores per notation method during minutes 0-5



Graph 4. number accuracy scores per notation method during minutes 5-10

The overall number accuracy scores are below:

Interpreter	1	2	3	4	5	6	average
score	38	28	40	38	26	40	35

Table 9. overall number accuracy score

5.5. Number accuracy differences between the two groups

The maximum achievable number accuracy score is 34 for the first half of speech and 16 for the second half. The score percentages for both halves of speech have been calculated from these maximum values and the accuracy values in Tables 7 and 8. The below tables include how many percents each participant achieved from the maximum number accuracy score in the respective segments of speech.

Interpreter	**1**	2	**3**	4	**5**	6	average
CASSIS used	**Yes**	No	**Yes**	No	**Yes**	No	
Minutes 0-5	**88,24%**	41,18%	**82,35%**	70,59%	**52,94%**	70,59%	67,65%

Table 10. number accuracy percentage during minutes 0-5

This means an average accuracy percentage of 74,51% with CASSIS and 60,78% in the control group **in the first 5 minutes** of the speech, amounting to an average Δ of **+14,27% of accuracy of numbers in favour of CASSIS**.

Participant	1	**2**	3	**4**	5	**6**	average
CASSIS used	No	**Yes**	No	**Yes**	No	**Yes**	
Minutes 5-10	50%	**87,5%**	75%	**87,5%**	50%	**100%**	75%

Table 11. number accuracy percentage during minutes 5-10

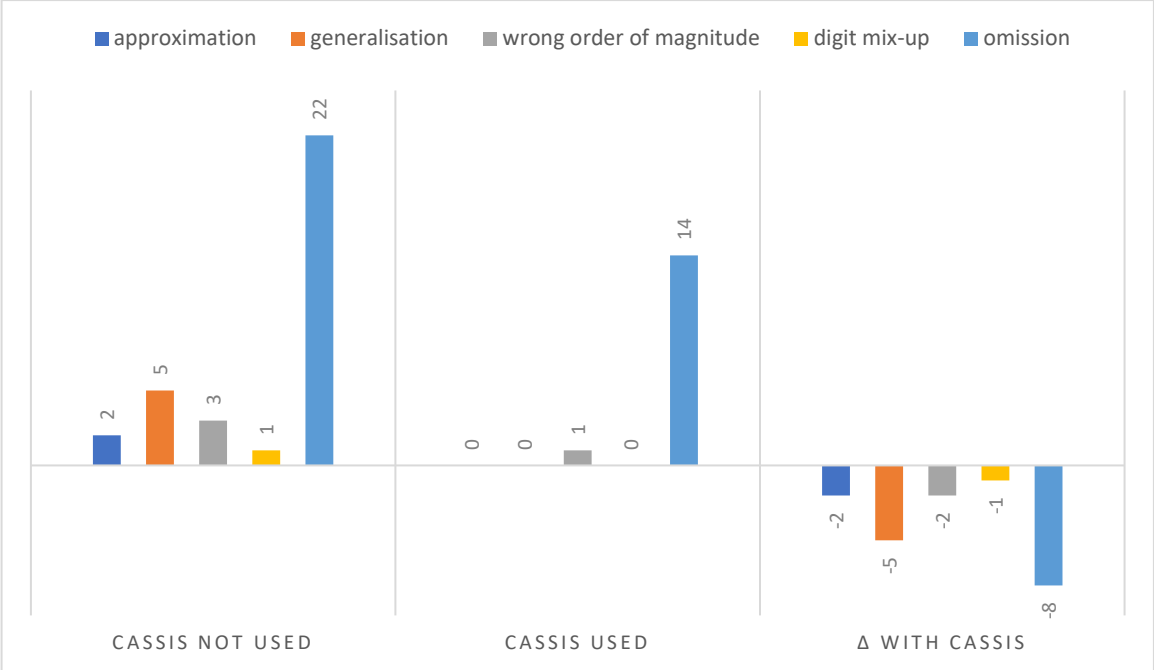
This means an average accuracy percentage of 90,67% with CASSIS and 58,33% in the control group on paper **during minutes 5 to 10** of the speech, amounting to a Δ of **+32,34% of number accuracy in favour of the CASSIS glossary.**

Each interpreter's number accuracy change can be calculated from the above two tables. The below values mean an average number accuracy Δ of +23,53% with CASSIS.

Interpreter	1	2	3	4	5	6	average
Number accuracy percentage with CASSIS	Higher	Higher	Higher	Higher	Higher	Higher	Higher
Δ numbers accuracy percentage	38,24%	46,32%	7,35%	16,91%	2,94%	29,41%	23,53%

Table 12. Δ number accuracy per interpreter

Different mistake types with numbers adapted from to Desmet et al. (2019) were observed: approximations, generalisations, using the wrong order of magnitude, and mixing up the first and second digits, as well as omissions were counted.



Graph 5. Mistake types when interpreting numbers

The rendition accuracy of different types of numbers was also summarised. Using CASSIS resulted in omission of 11 simple numbers, 2 years and 1 decimal number, furthermore 1 simple number in the wrong order of magnitude.

Not using CASSIS resulted in omission of 14 simple, 2 complex, 4 years and 2 decimal numbers. One simple, complex, and decimal number were translated in the wrong order of magnitude. 2 simple numbers were translated approximately. 1 simple number was reversed. 3 simple and 2 decimal numbers were generalised.

5.6. Enumeration accuracy scores

Evaluating the terminology accuracy scores of the 27 component terms of enumerations based on the scoring system of Prandi (2019), participants received 2 points for *close renditions*, 1 point for *acceptable renditions*, and 0 for *zero/unacceptable renditions.*

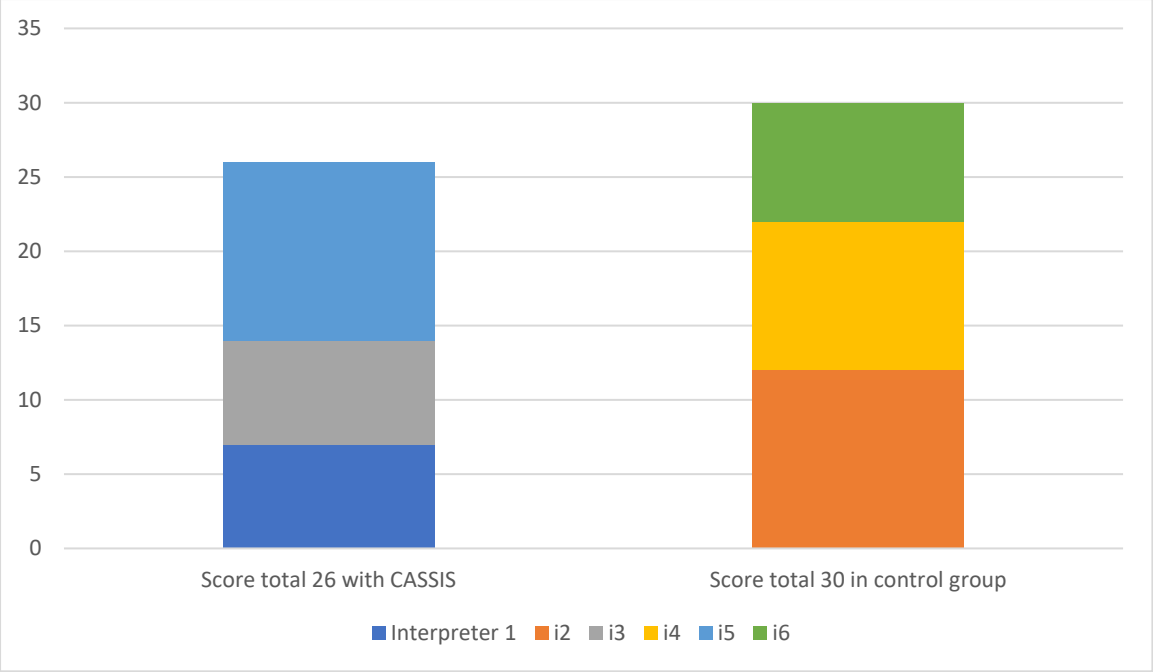
Similarly, compound terms which were enumeration components were given maximum 2 points each. For example: the enumeration "*from labour market reforms in Spain to pension reforms in Slovenia or to tax reforms in Austria*" includes the 3 underlined compound political terms regarding different reforms, each for a maximum accuracy score of 2. A close interpretation of the entire enumeration such as "*von Arbeitsmarktreformen in Spanien bis hin zu Rentenreformen in Slowenien oder Steuerreformen in Österreich*" amounted to a maximum of 6 accuracy points. Since only specialised terminology and numbers were given an accuracy score in this experiment, generalising the county names such as *"Arbeitsmarktreformen, Rentenreformen und Steuerreformen in der EU"* would also have resulted in 6 points for terminological accuracy. Saying "*Reforme in der EU*" would have resulted in 1 accuracy point, since it is a generalisation mentioning only reforms without specifics.

Participant	**1**	2	**3**	4	**5**	6	Average
CASSIS used min 0-5	**Yes**	No	**Yes**	No	**Yes**	No	-
accuracy score	**7**	12	**7**	10	**12**	8	9,33
accuracy percentage	**50%**	85,71%	**50%**	71,43%	**85,71%**	57,14%	66,67%

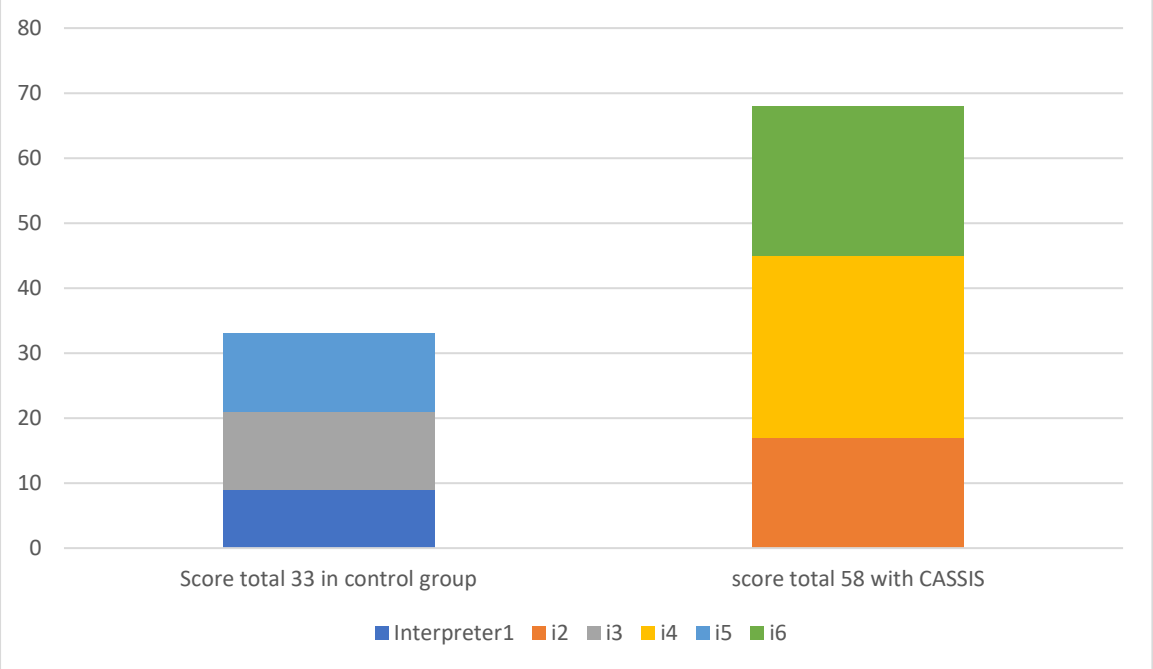
Table 13. enumeration accuracy scores minutes 0-5

Participant	1	**2**	3	**4**	5	**6**	Average
CASSIS used min. 5-10	No	**Yes**	No	**Yes**	No	**Yes**	**-**
accuracy score	9	**17**	12	**28**	12	**23**	16,83
accuracy percentage	22,5%	**42,5%**	30%	**70%**	30%	**57,5%**	42,08%

Table 14. enumeration accuracy scores minutes 5-10



Graph 6. enumeration accuracy scores per group from minutes 0-5



Graph 7. enumeration accuracy scores per group during minutes 5-10

We can calculate the change in accuracy of enumeration components from the percentages in the above tables 13 and 14. An average **+9,82% increase in accuracy of enumeration components when using CASSIS** was calculated.

Interpreter	1	2	3	4	5	6	Average
accuracy with CASSIS	Higher	Lower	Higher	Lower	Higher	Higher	Higher

Δ enumeration components accuracy	27,5%	-43,21%	20%	-1,43%	55,71%	0,36%	9,82%

Table 16. Δ accuracy of enumeration components

5.7. Coping tactics

The following coping tactics were recorded after analysing the interpretations:

Interpreter	**1**	2	**3**	4	**5**	6	Sum	
CASSIS used in minutes 0-5	**Yes**	No	**Yes**	No	**Yes**	No	**With CASSIS**	Without CASSIS
Delaying	**5**	2	**1**	1	**4**	1	**10**	4
Changing the ear-voice span	-	2	**1**	1	**6**	-	**7**	3
Segmentation	-	1	**1**	-	-	1	**1**	2
Generalising	**5**	3	**4**	4	**2**	-	**11**	7
Paraphrasing	**3**	5	**1**	5	**2**	1	**6**	11
Omission	**4**	5	**2**	6	**8**	2	**14**	13
Naturalisation	2	-	1	-	2	-	**5**	-
Transcoding	**1**	-	**1**	-	**-**	-	**2**	-
Reconstructing from context	**1**	2	**-**	-	**-**	-	**1**	2
Changing the order of elements in an enumeration	**-**	-	**-**	1	**-**	-	**-**	1

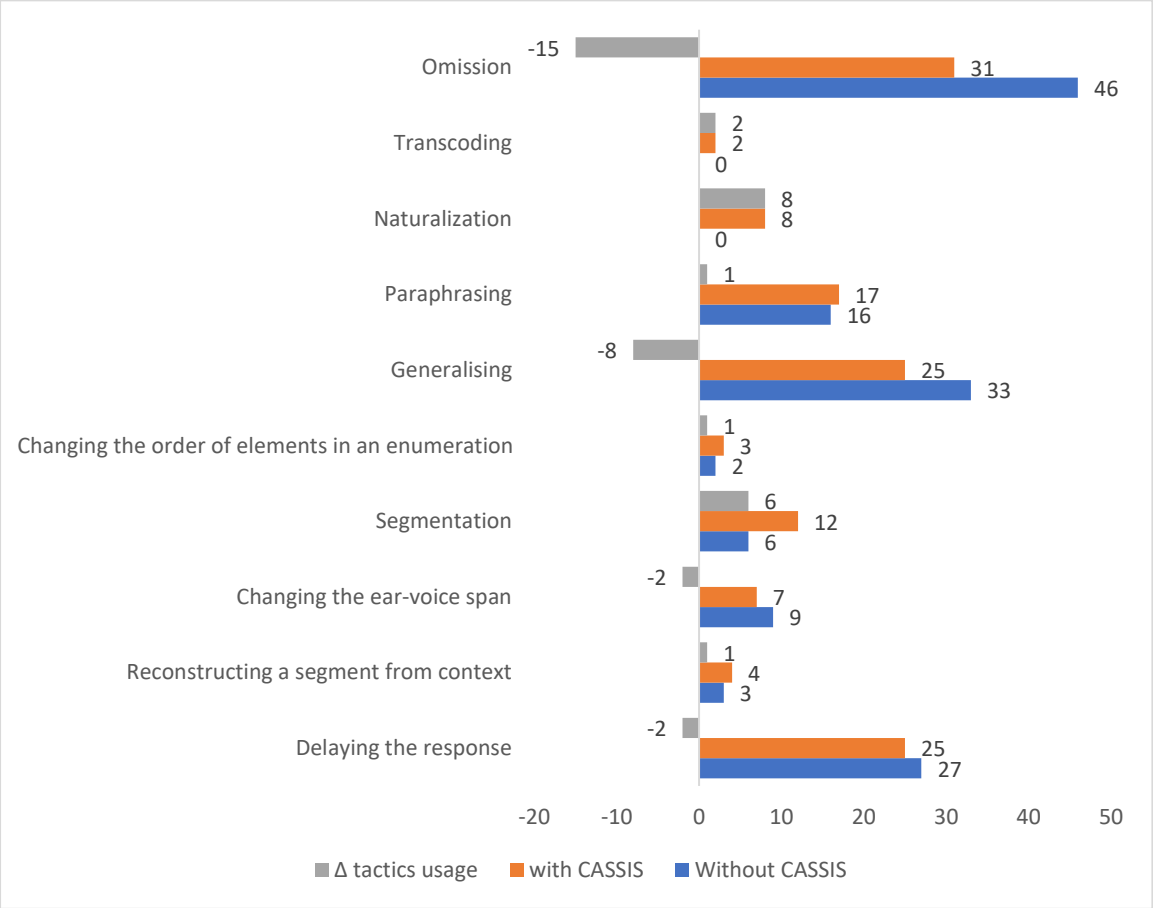
Table 13. coping tactics used on the selected specialised terminology minutes 0-5

Interpreter	1	2	3	4	5	6	Sum	
CASSIS used in minutes 5-10	No	**Yes**	No	**Yes**	No	**Yes**	**With CASSIS**	Without CASSIS
Delaying	2	**1**	1	**2**	3	**-**	**3**	6
Changing the ear-voice span	-	-	1	-	7	**1**	**1**	8
Segmentation	1	**2**	-	-	-	-	**-**	1

Generalising	5	**4**	4	**6**	7	**2**	**12**	16
Paraphrasing	6	**2**	2	**1**	-	**2**	**5**	8
Omission	7	**5**	8	**8**	18	**4**	**17**	33
Naturalisation	-	**1**	-	**-**	-	**2**	**3**	-
Transcoding	-	**-**	-	**-**	-	**-**	-	-
Reconstructing from context	1	**-**	-	**-**	-	**3**	**3**	1
Changing the order of elements in an enumeration	1	**-**	-	**2**	-	**1**	**3**	1

Table 14. coping tactics used on the selected specialised terminology during minutes 5-10

The changes in coping tactics usage when encountering specialised vocabulary can be calculated from the values in table 13 and 14. A positive Δ (delta) means the tactic was used more with CASSIS than without, and a negative Δ means it was used less with CASSIS than without it.



Graph 8. coping tactics for specialised terminology

The following coping tactics were estimated after analysis of the recordings and interviews:

Interpreter	**1**	2	**3**	4	**5**	6	Sum	
CASSIS used in minutes 0-5	**Yes**	No	**Yes**	No	**Yes**	No	**With CASSIS**	Without CASSIS
Consulting resources in the booth	**9**	1	**10**	-	**6**	-	**25**	1
Taking notes	**-**	-	**-**	-	**-**	3	**-**	3
CASSIS used in minutes 5-10	No	**Yes**	No	**Yes**	No	**Yes**	**With CASSIS**	Without CASSIS
Consulting resources in the booth	4	**7**	2	**6**	-	**7**	**20**	6
Taking notes	-	**-**	2	**-**	1	**-**	**-**	3

Table 15. estimated use of resources in the booth

5.8. Data concerning external factors

Some factors that might have affected the results have been included in the questionnaire: background knowledge, interpreting experience, preparation, and language combination.

Looking at language combinations, the lowest terminology score was achieved by participant 5 who interpreted from language C to B. The highest was achieved by no. 6 interpreting from B to A.

The participant with the lowest score also had the least simultaneous experience since his studies were focused on consecutive interpreting.

Since testing participant's preparation and background knowledge is beyond the scope of this paper, participants were asked in the questionnaire to assess their own preparedness and background knowledge once before and once after interpreting from 1 to 10. The below table will show the average of self-assessments in topics *politics, medicine/health, technology, economics* before and after interpreting. The table also shows the change in self-assessment after the assignment.

The questionnaire also measured how much time they have spent preparing for this one interpretation which will be shown in the table below.

Interpreter	**5**	**1**	**3**	**4**	**2**	**6**
Terminology accuracy score	**62**	**81**	**86**	**98**	**116**	**136**
Language combination used in this study	C->B	C->B	A->B	C->B	C->B	B->A
Interpreting experience	student*	student	student	0-5 years	0-5 years	student
Studies	Advanced MA*	Advanced MA	Advanced MA	MA degree	MA degree	Advanced MA
Time spent on preparation	<0,5 hours	0,5- 1 hour	<0,5 hours	<0,5 hours	1 – 2 hours	0,5- 1 hour
background knowledge self-assessment before	6	6,25	5,5	6,25	3,5	5
background knowledge self-assessment after	4,75	6	4,75	6,25	4	6
Background knowledge Self-assessment adjusted by	-1,25	-0,25	-0,75	0	+0,5	+1
cognitive capacity self-evaluation pre	5	7	7	9	5	7
cognitive capacity self-evaluation post	7	6	7	8	3	6
frequency of looking at CASSIS	10	8	7	9	8	9

Δ term. accuracy with CASSIS	11,19%	20,07%	15,17%	-5,53%	-4,81%	6,32%
Enumeration score	24	16	19	38	29	31
Δ enumeration accuracy with CASSIS	55,71%	27,50%	20%	-1,43%	-43,21%	0,36%
Number accuracy score	26	38	40	38	28	40
Δ number accuracy percentage	2,94%	38,24%	7,35%	16,91 %	46,32%	29,41%

Table 16. data concerning external factors

* Studies consecutive interpreting

The lowest terminological accuracy score, 62, was achieved by the participant with the least amount of simultaneous experience who studies consecutive interpreting with only an entry level simultaneous course.

The highest terminological accuracy score, 136, was achieved by the only participant interpreting into their A language.

Out of those interpreting into their B language, the two highest terminology accuracies, 98 and 118, were achieved by the two participants who have professional interpreting experience.

The two interpreters with professional interpreting experience have lost in terminological accuracy when interpreting with CASSIS. This happened both in terms of overall terminology, by -5,53% and -4,81%, as well as in components of enumerations by -1,43% and -43,21%.

The three lower scoring participants in terms of terminology (scores 62, 81, 86) have all decreased their background knowledge self-assessment after interpreting. One can also observe that the three participants who scored lower overall in terms of terminological accuracy, have made the larger gains in terminological accuracy, between +11% and +21% using CASSIS. This can be because if interpreters are scoring lower, they have more potential to improve their interpretation.

It is possible that the lower scoring interpreters improved their accuracy scores due to this effect. Moreover, all three of them were interpreting minutes 0:00-05:00 with CASSIS. Data suggests that terminology in this segment was easier to interpret correctly, which might have played a role in their larger gains with CASSIS as well.

The three higher scoring participants (scores 98 ,116, 136) have either increased or did not change their background knowledge self-assessment afterwards. The three participants who have scored higher overall for terminological accuracy, have all gained less than 7% with CASSIS, and two of them have decreased their terminological accuracy by between -4% and -6% using CASSIS. All three of them have interpreted minutes 05:05 -10:05 with CASSIS. In this half of the text, average terminological accuracy across all interpreters declined by 8,4% compared to the first half of the interpretation. Factoring in this average decrease in the second speech half, the three higher scoring participants still did not gain as much accuracy from using CASSIS as the three lower scoring ones.

The lowest scoring interpreter in terms of accuracy has achieved the lowest increase, 2,94%, in accuracy of numbers when using CASSIS. Coincidentally, they were the participant who completed only an entry level simultaneous interpretation practice and whose studies were focused on consecutive interpretation.

No further correlations could be assessed from this data in terms of number accuracy, since all participants have made measurable improvements of between 7% and 41% with CASSIS regardless of their overall score, which half of text they were interpreting with CASSIS or their interpreting experience or other external factors.

An attempt was made to study how this experiment affected the available cognitive capacities of interpreters, i.e., how much cognitive overload was caused by this interpretation. Participants have answered how mentally exhausted they feel after the experiment compared to before. The answers did not correlate with their overall score, which half of text they were interpreting with CASSIS or their interpreting experience or other external factors. Interpreter 5 even reported feeling they have better cognitive capacities after the experiment, after having looked at the software very frequently during interpreting. A verbal memory test was also conducted with participants, however the results proved too inconsistent. Some participants seemed not to take the online verbal memory test seriously achieving only a score of 10 and

some others achieved over 100. There was no correlation found between the verbal memory test results and any of the other study data.

6. Qualitative study data

6.1. Interviews

Study participants were asked to compare advantages and disadvantages of the different glossary methods respectively.

The number recognizing feature of CASSIS was reviewed most positively since 5 out of 6 interpreters (i1,2,3,4, and 6) found it helpful. Interpreters 1,2, and 6 mentioned the term base matches as a positive feature and found it helpful as long as the underlying term base is written well. Interpreters 2,3,4, and 6 found the software helpful when their time lag increased. Interpreter 1 mentioned it helped them when they had difficulty recalling a term from memory, but that they preferred only to look at the transcription if encountering a difficult sentence or term. Interpreter 3 reported "even if there were instances where I felt like it distracted me, I could easily look away and just concentrate on listening to the source speech". Interpreter 2 reported a boost in confidence knowing that there is a CAI tool they can use for help with difficult passages. The adjustable font size and the good visualisation of content were mentioned positively by one participant each.

Four interpreters (i2,3,4, and 6) mentioned mistakes in the live transcription of CASSIS as something that caused confusion during their interpreting, while one of them mentioned they saw words "disappearing" from the live transcription. Interpreter 2 mentioned the live transcription "corrects itself on the go" which is how this type of ASR software works. Interpreters 4 and 5 mentioned that the transcription was "slow", meaning it is not immediately transcribing the source speech as the underlying ASR software is processing the input. Interpreter 2 mentioned that term base matches were only shown once – if the same specialised term appeared again in the text, it was highlighted in the live transcription but its match from the term base was not written out again.

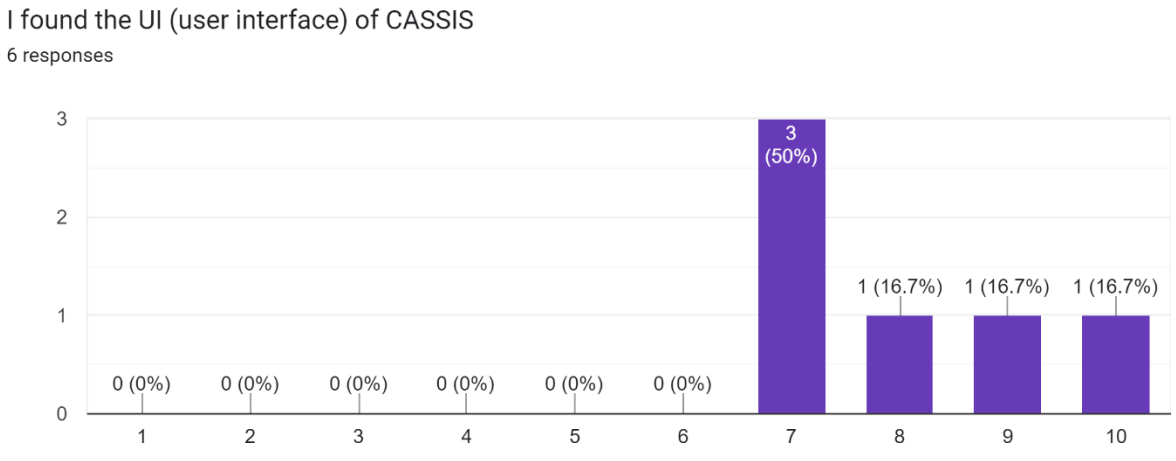
When asked to describe using the spreadsheet glossary, 3 out of 6 interpreters (i1,3, and 6) said that it took too much time to look up words during SI from their printout or Excel glossary and doing so "caused distractions". Interpreter 2

reported not having paid attention to “hard words” because of the time it takes to find them using the spreadsheet glossary, but rather having used coping tactics to get around them.

Interpreters 2,3, and 6 reported that the traditional glossary was useful during the preparation phase of an assignment, and it helped them memorise terminology beforehand. Interpreters 4 and 5 expressed that they trust their usual glossary method more because they already know which term is where in their glossary and translated terms are accessible in the paper glossary all the time. Interpreter 1 said it is easier to take to the booth and there is no software setup required.

When asked how they would improve the functionalities of the CASSIS software, 2 out of 6 interpreters (i1, and 3) mentioned they would like numbers and units of measurement to show up separately, because it “can be hard to find numbers that are spelled out instead of written with digits”. Two interpreters provided suggestions for improving the live transcription, however it is unclear whether these are implementable in the software: interpreter one suggested an error-tolerant term base search, and interpreter 5 said words not matching the context could be separated within the live transcription. One suggestion was made by interpreter 2 to write out term base matches every time a certain term is recognised in the transcription. Interpreter 6 missed looking at the video of the speaker and recommended adding the video (in remote SI mode) to the CASSIS interface.

The user interface was evaluated on a scale from 1-10 where 1 is hard to understand and 10 is easy to understand:



Study participants received a user manual for CASSIS, however, they did not have to set up CASSIS themselves. When they began the experiment, the term base was already uploaded into CASSIS, the input language and the correct audio input settings were applied, and the font size enlarged to 16 points to make it easier to read. Therefore, this score only applies to the UI during the interpretation phase of an assignment, but not to the preparation phase where the aforementioned steps must be carried out.

Interpreter no. 2. and 4. reported in the interview that when they were looking for terminology from the automatic term base, they were sometimes looking “at the words highlighted in yellow”. This could cause problems, since terminology is only highlighted in the ASR transcription, which is in the source language. The suggested translations on the right-hand side column of the CASSIS screen are not highlighted in any colour. If the yellow highlight draws attention away from the suggested translations, that could defeat the purpose of those translations. Below is an illustration of the highlights in the source speech transcription and the lack of highlights in the suggested TL terminology:

ASR transcription	Term base matches
The aim is to build a consensus well ahead of 2023	to build a consensus → einen Konsens erzielen

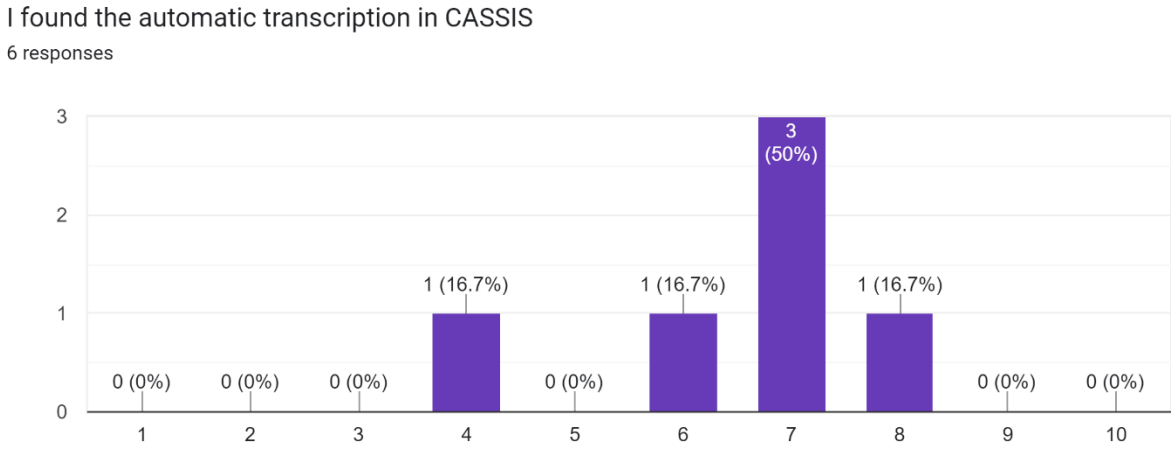
Table. 17. Illustration of highlights

Crucially, interpreters 2 and 4 both had a slightly lower terminological accuracy when they had access to the CASSIS screen. Looking at the highlighted words in the source language transcription could have influenced that, since then they would have to switch their attention over to the other side of the screen to find the same source language word, and then the German equivalent from the term base. Looking for the highlighted words frequently means also frequently looking at the automatic transcription, which with its mistakes reportedly lead to further distractions.

6.2. Feedback regarding the automatic transcription

Feedback from the interviews was cross-checked with the recorded interpretations. Deductions in the following chapter were made based on the recordings and the interviews made during the experiment but were not verified with eye tracking technology.

The automatic transcription feature was rated on a scale of 1-10 where 1 means not useful and 10 means very useful:



A common pattern when using the transcription feature of CASSIS was reported by 3 out of 6 participants: Interpreters 2, 3, and 4 read the live transcription at the start of the interpreting. However, all 3 of them reported that after some issues they encountered in the transcription they did not read along with it anymore because of a loss of trust.

Interpreter 2 was distracted by the fact that the transcript "corrects itself on the go". For example, the address at the beginning "*Honourable Members!*" is spelt out in several steps. These versions are *under -> honour -> honourable -> honourable men -> honourable member -> honourable members.* The transitions between the aforementioned versions take up only a few frames of the recorded video, and the human eye can only read all of these steps if the recording is paused frame by frame. In real time, one can only see the text changing very rapidly, then changing into its result after less than 1 second. However, for interpreter 2, the fast change was still distracting enough to make them report a fault in the live transcription.

Interpreter 3 reported that they started thinking what a mistakenly spelt word meant. The term "*European Health Union"* [8], was misspelt as "*European house union*", which audibly cost them time in their interpretation and caused a pause and "*erms*" from 03:20 until 03:23 in the recording. After these 3 seconds, the interpreter opted for a gross generalisation and used the German equivalent of "*result*" – "*Ergebnis*". This was evaluated as an unacceptable rendition for 0 points. Their terminology accuracy

[8] A European Health Union was mentioned in the source speech as a theoretical concept

was 78,5% up until this mistake at 03:20 and 47% afterwards until the changeover at 05:00. However, it is unclear whether this was caused by them not looking at the transcription anymore, or the text being more difficult after 03:20, or mental fatigue or some other external factor. Their terminology accuracy in the second text half using a spreadsheet glossary was 38,24%, which is still lower than after the mistake at 03:20 with CASSIS. This correlates with them reporting that CASSIS was overall helpful for them, but they found looking up words from the spreadsheet glossary after minute 5 did not help their interpretation as it took too long to find terms that way. However, they also felt that their background knowledge and preparation was better for the topics in the first half than in the second half, so this and other external factors may also have contributed to the decline in their terminological accuracy score after minute 5.

Interpreter 5 mentioned that they found the speed of the transcription slower than they were expecting, however they still reported having looked at CASSIS frequently. This participant expected the transcription to function similarly to movie subtitles: They thought CASSIS was going to show the transcription and the matches along with the source speech, because they were not familiar with how an ASR system works. Therefore, this person was disappointed by the up to 2 seconds of delay, even though his ear-to-voice span during interpretation was longer than 2 seconds. This ear-to-voice span means that this participant has indeed had the possibility to use the term base matches and the live transcription and that their comment was rather motivated by their own personal expectations.

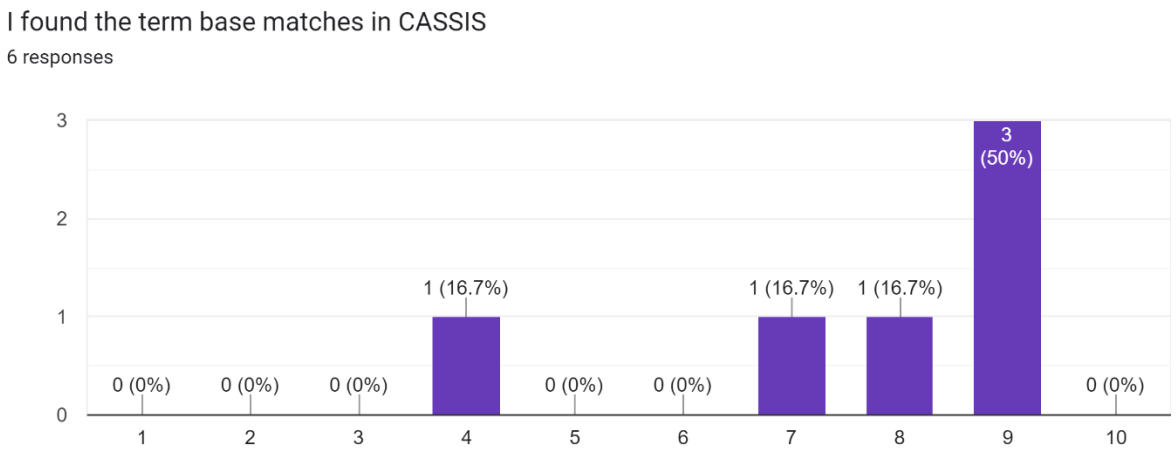
It is possible that interpreter expectations for the automatic transcription were too high. None of the participants were informed before the experiment about the way the ASR system works. Namely, that it is not instantaneous but can have a lag of up to 2 seconds. An automatic speech recognition software needs time to find the correct transcription of uttered words. The CAI tool then searches through the uploaded term base for matches. Therefore, it took a delay of approximately 0,5 seconds up to 2 seconds from when a word was uttered until a term base match in German was shown. The ASR transcription had a word error rate of 7,4% in the first half and 7,5% in the second half of speech, higher than the recommended value for interpretation (Fantinuoli 2017). This is where some of the reported frustration might have originated. However, even after some interpreters said they stopped following the live

transcription, they reportedly still looked at it for numbers and looked for matches from the term base.

In CASSIS, the numbers are only written out in the automatic transcription, and they are not highlighted nor are they written out in a separate column. Two interpreters have mentioned that they would like this feature to be improved. However, even these two agreed that seeing the numbers in the text was helpful and the study data shows an improvement in number accuracy when using CASSIS for every participant. Only interpreter no. 5 and no. 6 reported that they noted numbers on paper along with their traditional glossary. The other 4 participants interpreted numbers only relying on their short-term memory when not using CASSIS. There were no boothmates during the experiment. This means, the data does not show if using CASSIS yields better results than noting down numbers or having a boothmate write them down.

6.3. Feedback regarding the term base matches

The term base match feature of CASSIS was rated on a scale of 1-10 where 1 means not useful and 10 means very useful:



A weakness of the term base matches originated in the underlying ASR system. If a term was in the bilingual term base and in the source speech as well, however it was not recognised by the automatic speech recognition system, it was never shown during the experiment. Participant 6 reported to look for the German equivalent of the concept "*European tech sovereignty".* However, the term was misrecognised by the ASR as European *text* sovereignty and the study participant misheard it as European *tax* sovereignty. Since the ASR did not transcribe the English

term “*European tech sovereignty”,* CASSIS has not searched though the term base for this term, and the German equivalent was never suggested in the right column of the screen. Even though neither *tax* nor *text* fit the context well, this interpreter has interpreted what she has misheard as “*European tax sovereignity*” into “*Europäische steuerliche Unabhängigkeit*”. In this case, neither the ASR nor the interpreter has deduced the correct source language term from the context. The question is, if the participant could have prepared better to overcome this difficulty of if they relied too much on the German term being shown automatically. Additionally, none of the participants interpreted European tech sovereignty, regardless of glossary method. That means, this misunderstanding might have been caused by other factors as well, for example the accent of the speaker or the overall difficulty of the source speech.

The same interpreter reported that sometimes the verbs that they could not recall from memory were not shown by CASSIS in German. This depends on the underlying term base. Only verbs that were specialised terminology were included in the term base used for this experiment. However, there is a possibility to add verbs to the term base in CASSIS. Interpreter 6 also recalled that the term base matches sometimes helped with memory issues: They recalled one occasion looking for the German equivalent of “*competitiveness*” and incorporating the match from CASSIS “*Wettbewerbsfähigkeit*” into their interpretation.

Interpreter 5 has not found the automatic term base matches useful, however their recorded interpretation does not entirely match their report. At 3:40 in the source speech there is a passage that this participant possibly interpreted using the term base matches from CASSIS. Specialised terminology that was included in the term base and shown along with their German equivalent on the CASSIS screen during interpretation are underlined in the following passage: *We have the innovation and scientific capacity, we have the private sector knowledge, we have competent national authorities, and now we have to bring all of that together including massive funding.* The interpreter has paused briefly at the start of this long enumeration, then continued to incorporate all term base matches shown on the right-side column of CASSIS into his rendition. The terms this study participant used from CASSIS are underlined below in the transcript of their rendition. *Wir haben die Innovation, die wissenschaftliche Kapazität, * wir haben die/ das Wissen des Privatsektors, * und wir haben die zuständigen nationalen Behörden. /auch/ * Und dadurch können wir * massive*

Finanzierungen vorbereiten. Even though the interpreter did not specifically mention reading the terms from the automatic glossary, the quoted sentence includes all terms from the screen. This is in contrast with the rest of their interpretation where sentences were very frequently generalised, and many details were omitted.

Interpreter 3 reported an issue with the term base matches, but their recollection of events was proven as a false memory by analysing the recording. The interpreter had an ear-to-voice span of about 2 seconds. They remembered looking for the German equivalent of the "*single market"*, but it not yet being shown up. To check their account of this event, the interpretation can be cross-referenced with the captured screen of CASSIS. From the time where the term "*single market*" was uttered, approximately 1 seconds has passed until the match "*Binnenmarkt*" from the term base was shown up on screen. So even though the source term was highlighted in the transcription, and the match was shown up, this interpreter has failed to notice it for some reason and ended up generalising it as "*Markt*". However, the interpreter could not recall why they missed the term. This could have been due to an overload in cognitive capacities, or that the German word was shown sooner than they expected and therefore they missed it. It could have been caused by the highlighting issue as well, where only source language terms are highlighted but not their suggested target language equivalents.

6.4. Feedback about the number recognition feature

The number recognition feature found the best acceptance by study participants. All of them reported having read numbers from the ASR transcription, and 5 out of 6 found it useful. This does not mean that the other interpreter found it useless, rather that they felt this feature could be improved by writing numbers with digits instead of sometimes spelling them out in the transcription.

The overall approval of this feature correlates to the study data since all 6 interpreters improved their interpretation accuracy of numbers. Approximations, generalisations, and digit mix-ups were eliminated using CASSIS, and only one mistake was made interpreting the wrong order of magnitude. This is a result of "*one billion*" being equal to "*eine Milliarde*" and not "*eine Billion*" in German which made the interpreter mix it up, even though they saw the transcription in English.

6.5. Patterns in interpretation

6.5.1. In-depth analysis of looking up terms

To provide an example of looking up terms from CASSIS compared to looking them up from a glossary sheet, the interpretations of the following two sentences at 07:20 in the source speech will be analysed: "*And, to do that, the Commission will relaunch the discussion on the Economic Governance Review in the coming weeks. The aim is to build a consensus on the way forward well ahead of 2023*." The underlined terms were in the bilingual word list sent to the interpreters ahead of the experiment, as well as in the bilingual term list in CASSIS. The term base matches were all displayed by CASSIS. Even though the term Commission was highlighted in the left side transcription, it was not displayed on the right side with its suggested German translation, because the same term appeared earlier in the text and CASSIS displays the suggested translation only the first time.

Interpretations of these two sentences by all interpreters are transcribed and analysed in this chapter. Self-corrections in the German interpretations are marked with a slash, audible hesitation sounds transcribed as "*öhm*", pauses of up to half second marked with *, up to one second marked with **, more than 1 second marked with (*pause duration) and the German term from the bilingual glossary marked with a dotted line. Interpreters 1, 3 and 5 had no access to the CASSIS screen during this speech segment while interpreters 2, 4 and 6 has CASSIS on their screen.

Interpreter 1 reported that they did not recall the German term for *Economic Governance Review* from memory, but they remembered reading it on the bilingual term list, therefore they decided to look it up from their Excel spreadsheet. Their German interpretation was: "*Dazu m/ * muss/ * wird öhm die ** öhm * in den nächsten Wochen eine neue Diskussion über die Überprüfung der wirtschaftspolitischen Steuerung geben*." These delays were likely used by the interpreter to look up the German term "*Überprüfung der wirtschaftspolitischen Steuerung*". During the time that they looked up the term, they omitted the actor: "*the Commission"*. Furthermore, the next sentence: "*The aim is to build a consensus on the way forward well ahead of 2023*" was missing altogether from their interpretation and they immediately continued interpreting the sentence after that. This is an example of how much time and effort it takes to look up a term during SI from an Excel glossary. The combined time of all

audible self-corrections, pauses and hesitation sounds until the interpretation continued fluently, was about 7 seconds in total.

Interpreter no. 3. did not have access to the CASSIS screen at this segment and decided to look at the bilingual glossary from time to time, but it reportedly did not help. However, it was unclear whether they looked at their bilingual term list during these two sentences, or if they were just pausing for another reason, perhaps changing the ear-voice span to gather more information from the source sentence. Their interpretation is transcribed here: „*Und die/ das/ die Kommission wird ** eine Diskussion führen über die ** öhm * darüber, und Ziel ist es, dass wir einen Konsens erzielen vor dem Jahr 2023.*“ The dotted underline marks a German equivalent for “*to reach a consensus*” using a different verb than was suggested in the bilingual glossary. Therefore, it is likely that the interpreter did not look up this expression from the glossary, and instead recalled it from memory. The used verb “*erzielen*” is still a correct interpretation. The subject of the sentence “*the Commission*” and the year are also interpreted correctly. This interpreter, however, missed the German equivalent of “*the Economic Governance Review”* and substituted it with the adverb “*darüber*” after pauses and hesitation sounds of about 5 seconds in total. This adverb points to the sentence before, which was not about the Economic Governance Review but rather about the effects of the current crisis. This in turn would cause listeners to potentially misunderstand the speaker’s message.

Interpreter 5 had no access to CASSIS but reported having used the spreadsheet glossary occasionally. Their interpretation is transcribed here. “*Diese Kommission wird diese Diskussion (*2s) in den folgenden Wochen * einen Konsens bauen ** öhm, bis 2030*“. In this interpretation, *"the Commission”* and “*to build a consensus”* are both interpreted adequately, however the sentence is incomplete, and details are missing. This interpreter reported that the spreadsheet helped them prepare vocabulary for this assignment, however they could not recall the key object “*the Economic Governance Review”* and there is an audible pause of approximately 2 seconds at its place. This might have been due to issues with recollecting the term from long term memory, or simply due to an overload on mental capacities.

Interpreter no. 2 had access to the CASSIS screen during this passage. Their interpretation was: „*Und auch die Kommission muss das diskutieren. Es ist, öhm, sehr*

*wichtig, * diese Überprüfung der wirtschaftspolitischen Steuerung. Wir müssen zu einem Konsensus gelangen, öhm ** bis 2023.*“ This interpretation correctly included specialized terminology from the source sentences. There is also an audible pause and hesitation sounds ahead of the German term for *the Economic Governance Review* as well as ahead of the year 2023, possibly caused by the interpreter looking them up from the CASSIS. This interpreter reported to first having read the terms highlighted in yellow in the full transcript on the left side column, then looking at the suggested German translation on the right-side column. Notably, the interpreter used the German word “*Konsensus*” which is phonetically close to the English term “*consensus*” in the source speech, however it is not frequently used. The German term “*Konsens*” is more widely used. This might have been a language interference caused by following the ASR transcription. Even though terminologically, the interpreted segment was complete, the interpreter’s choice of auxiliary verbs is not entirely accurate. He used the German verb “*müssen*” which points to a necessity, rather than plans and aims which were expressed in the source speech. So even though this interpreter received 6 out of 6 points in the analysis of terminological accuracy and 2 out of 2 points for number accuracy, the auxiliary verbs could have been chosen more accurately.

Interpreter no. 4. had access to CASSIS for this segment: “*Wir müssen in der Kommission daran arbeiten, dass in der Wirtschaft mehr getan wird, damit es ein/ damit ein Konsensus erreicht wird, damit wir uns weiterbewegen, damit wir das erreichen bis zum Jahr 2023, worüber wir gesprochen haben*.“ This is a very fluent interpretation with no hesitation sounds, no breaks, and only one self-correction. The interpreter mentions the economy: “*Wirtschaft*” then interprets the expression “*to reach a consensus*”. There is a phrase marked in grey that can be translated back into English as “*so that we move forward*” which was not in the two source sentences, rather used as a coping tactic to delay the response and possibly to circle back to the missed terminology. Despite this delaying tactic, interpreter 4 missed the German equivalent of “*Economic Governance Review*” and used a very broad generalization instead. The listeners could only understand that the economy will be discussed. It is unclear why the term displayed by CASSIS was missed in this case. The same interpreter reported in the interview another occurrence, where they failed to spot a suggested German translation for the “*European single market”*, but the analysis of

the recording revealed that the term was there on the CASSIS screen. Therefore, it is possible that this interpreter had a difficulty in understanding the layout of the CASSIS screen and where to look for translations.

Interpreter no. 6. was interpreting into their A language, therefore their rendition can be expected to be more fluent. At the same place at 07:20 in the source text, interpreter no.6. correctly interpreted all three terms automatically suggested by CASSIS screen: „*Die Kommission wird daher * Diskussionen starten, zur Überprüfung der wirtschaftspolitischen Steuerung in den nächsten Wochen*, *und * wir wollen zu einem Konsens gelangen, und das vor 2023.“* The word order is irregular for German grammar since the infinite verb “starten” should come after the object “*Überprüfung der wirtschaftspolitischen Steuerung”* and the temporal indicator “*in den nächsten Wochen”.* However, there are only two pauses of about half a second, no hesitation sounds, and all details were interpreted from the two source sentences. The delay caused by looking up the terms in CASSIS was in this case about half a second, and it occurred two times. However, interpreter 6 had a complete and accurate interpretation. All terminology, auxiliary verbs are interpreted correctly, and the speaker’s intended message is completely transferred into German by using incorrect word order, segmenting the text, pausing briefly before incorporating the suggested translations and numbers.

6.5.2. Interpreting numbers

In this chapter, the differences in interpreting numbers will be analysed. This is not a comparison of using CASSIS with noting numbers, since only interpreter 6 decided to note numbers. The group that did not have access to CASSIS can be therefore evaluated as a control group. The analysed section of the source speech is at 04:33 in the source speech. “*Today, more than 400 million certificates have been generated across Europe. 42 countries in four different continents are plugged in.*” All interpreter renditions will be analysed here focusing on how they interpreted numbers. Self-corrections in the German interpretations are marked with a slash, audible hesitation sounds transcribed as “*öhm*”, pauses of up to half second marked with *, up to one second marked with **. Interpreters 1, 3 and 5 had access to CASSIS in this segment. Interpreter 6 reported to have noted numbers down, interpreter 2 and 4 have relied on their short-term memory for numbers.

Interpreter 1's rendition is transcribed here: "*Mehr als 400 Millionen öhm *, EU digitale Zertifikate wurden bis heute erstellt*." Since this interpreter had a time lag of about 10 seconds, they only summarized these sentences and missed 2 out of 3 numbers, even though they could have easily read them from the screen. Problems earlier in their interpretation have caused this incomplete interpretation. Notably the interpreter had a short pause and hesitation sounds after getting the number correct.

Interpreter 3 has recorded the following: "*Mehr als 400 Millionen Zertifikate wurden ** generiert, öhm, * über Europa, und 42 Länder in 4 verschiedenen Kontinenten * sind auch dabei.*" This interpreter transferred all numbers correctly into German. This interpreter also paused after the number 400 million.

Interpreter 5 transferred into German: "*Wir haben mehr als 400 Millionen Zertifikationen * generiert. * Mehr als 42 europäischen Länder * auf 4 Kontinenten*." Although all the numbers in this interpretation are correct, the second sentence is incomplete since there is no verb. If translated back into English, the second sentence sounded: "*More than 42 European countries * on 4 different continents* **. This contradiction shows that it is not enough to read the numbers from the CASSIS screen, one also must fit them into the interpretation.

Interpreter 2 has interpreted the sentence as follows: „*Das digitale Zertifikatssystem der EU. * Es sind mehr als, öhm, * 24* Länder, öhm, verbunden worden.*" We hear pauses and hesitations, possibly to recall the numbers from memory, however this might be caused by this interpreter's specific style as it happened in the previously analysed segment as well. They swapped the numbers 42 to 24 due to linguistic difference between English and German. This interpreter missed "*four continent*s" and "*400 million certificates*" possibly due to the long pauses.

Interpreter 4's target text for these sentences was the following. "*Zum Beispiel das Zertifikat, das wir erarbeitet haben. Über 400 Millionen Zertifikate wurden innerhalb von Europa * erstellt. 42 Länder auf zahlreichen Kontinenten.*" Two out of three numbers were interpreted correctly. The phrase "*four continents*" was generalised as "*numerous continents*" which was counted as an acceptable rendition, as four continents are a large portion of the whole. The message of the source text was fully transferred except the one missing number and another number that was

only approximately interpreted. This was an overall fluent interpretation with a grammatically incomplete sentence.

Interpreter 6 had the following Germen interpretation. "*Sehen wir zum Beispiel das digitale Zertifikatssystem der EU an. Wir haben mehr als 400 Millionen Zertifikate in Europa ausgestellt in 42 Ländern, in vier/. Auf vier verschiedenen * Kontinenten sind alle * verbunden*." This interpreter has interpreted all numbers in this sentence correctly and conveyed the message from the source text completely. This interpreter solved the contradicting switch from *Europe* to *4 different continents* as a scene by self-correcting and marking the start of a new sentence with their intonation. This way, the 4 different continents are in a new sentence and do not contradict Europe from the precious sentence.

Interpreters who did not see the automatic transcription during this segment have in total interpreted the numbers slightly less accurately. Even though interpreters with CASSIS on their screen saw all numbers, there was one incomplete interpretation due to a big delay caused by previous sentences of interpreter 1. And even though interpreter 5 transferred all the numbers correct, there was a logical mistake in their German interpretation which they could not correct. Interpreter 6 managed to correct this apparent contradiction by marking a new sentence with intonation. Interpreter 4 who did not have access to CASSIS and was not taking notes, has interpreted these two sentences completely by using just one generalization. This makes it evident that using coping tactics strategically can improve the interpretation even if not all numbers are transferred completely.

6.5.3. Interpreting when the transcription was faulty

Any automatic match from the term base depends on the automatic transcription to correctly identify that term. Below are some examples, where the transcription did not pick up a term in question, or transcribed it erroneously:

In the first example, all but one interpreter missed the term that the ASR system transcribed incorrectly. At 5:05 in the source text the speaker says: "*And we did a lot of things right. we moved fast to create SURE. This supported over 31 billion/ million workers (…)."* The transcription instead picked up: "*We did a lot of things right. Remove a stupid show Vista Portage of a 31 billion million workers (…)."* It did not pick

up the acronym SURE, nor its context correctly. Only interpreter 1, who was in the control group during this segment, interpreted this acronym. This interpretation is transcribed below. "*Und wir haben, öhm, viele Sachen richtig gemacht. Wir haben, öhm sehr schnell, öhm * die öhm * die/ das SURE-Instrument, öhm ins Leben gerufen. Das hat, öhm, mehr als 30 Millionen Arbeitnehmerinnen und Arbeitnehmer (…) unterstützt*." Possibly, this interpreter has interpreted it from their glossary since there are audible pauses and self-corrections before the term which could be time to look up the word. However, hesitation sounds were characteristic for this interpreter throughout the whole product. All other interpreters have omitted this acronym and continued their interpretation. Interpreter 2 reported that the above transcribed mistake in the transcription distracted them, although they reportedly kept looking at the transcription afterwards.

In the second example, a term was transcribed incorrectly, but some translators could deduce or paraphrase it from context. At 04:30 in the source text, the speaker said: "*Take the EU Digital Certificate*" The ASR system transcribed "*Fix it you did it so certificate (…)*" 2 out of 3 interpreters who had CASSIS on the screen during this passage had only interpreted the term "*Zertifikat*" which is the German equivalent of the correctly transcribed term "*certificate*". However, these two interpreters did not provide more context in German and thus interpreted incompletely. 1 interpreter with CASSIS on screen correctly interpreted the full term "*das EU digitale Zertifikat*" with approximately 1,5 second delay, possibly to listen for more context. This interpreter however had omissions in the upcoming sentence. 2 out of 3 interpreters who did not have CASSIS on screen paraphrased adequately and accurately "*das digitale Zertifikatssystem der EU*". The third interpreter in the control group used an explanation strategy: "*das Zertifikat, das wir erarbeitet haben*", which means a loss in accuracy, but due to the explanation and the context, the listener may be able to deduce which certificate is being mentioned.

6.5.4. Omission patterns

Omission of redundant or inappropriate information in the source speech can be used as a coping tactic. "*When information is omitted, it is not necessarily lost as far as the delegates are concerned – it may appear elsewhere or be already known to the delegates*." (Gile 2009:210). In this experiment, interpreters did not know the

context for the speech, nor did they have any documents that may have been available to the listening members of the European Parliament at the time. There was nothing in the speech that could have been interpreted as inappropriate. The only possibility of using omission as a deliberate coping tactic is if something within the speech was repeated. There was only one redundant information in the speech, where the speaker mentioned twice in two sentences that the European single market had been existing for 30 years. If one participant missed this, it is possible they did so deliberately to avoid repetition. Any other omissions must have a different cause since this was the only possible place to omit deliberately.

Omissions could also occur if the interpreter "did not have enough processing capacity available for the Listening and Analysis Effort when the speech segment carrying it was being uttered. They may also omit it because it disappears from short-term memory." (Gile 2009: 210) Such mental overloads are worth discussing within this paper, since one of the aims is to determine whether an automatic glossary can help with Listening and Analysis or Memory efforts.

A table of total omissions counted can be seen below. The numbers are in bold where the interpreters had the CASSIS screen available:

Speech segment	Interpreter 1	i2	i3	i4	i5	i6	With CASSIS	Without CASSIS
Minutes 0-5	**4**	5	**2**	6	**8**	2	**14**	13
Minutes 5-10	7	**5**	8	**8**	18	**4**	**17**	33

Table 18. Number of omissions

Only participant number 4 has omitted the repetition of the 30 years old European single market in the second speech segment. Therefore, all other omissions are most likely to have been caused by some sort of mental effort overload based on the theory by Gile.

The total number of omissions had no significant variance in the speech segment from 0:00 until 05:00 between CASSIS users and the control group. There were 14 omissions with CASSIS and 13 without it. There was however a significant difference in the speech segment from minutes 5 until 10. 17 omissions happened

while CASSIS was on screen. One of these omissions was possibly a conscious coping tactic but 16 were likely to have been caused by mental effort overloads. 33 omissions were observed in the renditions where interpreters did not have CASSIS on screen. None of them were made at the identified redundant part, therefore all 33 were likely caused by insufficient cognitive capacities for Listening and Analysis and/or Memory efforts. There seems to be some correlation between having CASSIS on screen and the number of omissions made during interpretation.

Only participant no. 3 had the same number of omissions in both 5-minute speech segments. All other interpreters had more omissions in the second segment from minutes 5 until 10; therefore, it is reasonable to assume that the second part was more challenging in terms of the Listening and Analysis and/or Memory efforts. However, since it is impossible to measure the usage of these efforts, it is also not quantifiable how much of an effect this had on the number of omissions. Another detail worth mentioning is that the largest increase in omissions in the segment from minutes 5 until 10 was made by participant no. 5 who had the least amount of SI experience. This increase in omissions is most likely to have been caused by several factors together: the usage of CASSIS, the preparation level, experience of participants as well as their usage of the different mental efforts.

6.5.5. Individual patterns

Interpretations were analysed to see if and how using this glossary tool has changed coping tactics used during speech passages that caused problems for interpreters. One must view this analysis in the context that the source speech topics were changing throughout from mainly medical-political in the minutes 0 to 5 to finance and technology politics in the minutes 5 to 10. If an interpreter was better prepared in one topic than the others, that might also have influenced the problems they encountered and coping tactics usage. That is why the self-evaluation about interpreter's preparedness in these topics can be used to aid this coping tactic analysis.

Interpreter 1 tended to paraphrase rather than generalise while using CASSIS and they had less omissions during this segment. This might have been also aided by better preparation for the topic in the first half. Their self-evaluation of their background knowledge was 8 out of 10 for medicine for the segment with CASSIS. Their time lag was very long in both segments, so this might be just a general problem which caused

omissions in both segments. They had more omissions and more unfinished sentences in the segment without CASSIS, and they tended to generalise rather than paraphrase, when encountering problems. However, their interpretation sounded more idiomatic without CASSIS.

Interpreter 2 had some errors and misunderstandings caused by transcription errors in CASSIS during the minutes 05:05-10:05. They read the matches but sometimes misunderstood the source sentence in the aftermath. However, they also misunderstood the source text occasionally without CASSIS in the first half. The number of omissions stayed 5 in both speech halves, but the number of generalisations decreased from 5 to 2 when this interpreter used CASSIS.

Interpreter 3 was accurate both in terms of terminology, numbers, and in terms of transferring the message of the speaker. There were some generalisations and omissions when the CASSIS transcript was not helpful from 0:00-05:00. However, there were more omissions, a larger lag, and more misunderstandings in the second speech segment without CASSIS. Both generalisation and omission increased from 2 each to 8 each in the second half. This could be due to the different topic as reported in the post-experiment interview, and as the self-evaluation in technology and economics were lowered after the experiment.

Interpreter 4 had better results in the control group, except in the accuracy of numbers. Terminology accuracy percentage with CASSIS has gone down, the interpreter omitted and generalised more. This interpreter seemingly had a hard time using the suggestions from CASSIS – they reported not to see a suggested German translation even though it was on the screen at the right time. This interpreter also reported to read the live transcription and the words marked in yellow, which are only marked in the source language. The suggested target language translations are not marked yellow which is maybe the reason that some of them were missed. The usage of coping tactics without CASSIS was seemingly more tactical and yielded better accuracy overall, which could be the result of the professional experience of this interpreter.

Interpreter 5 has generalised a lot of the source text with both glossary methods. They used CASSIS audibly to help with components of enumerations which gave them good accuracy with those terms. They have also improved their number accuracy with CASSIS.

Interpreter 6 had good effort management and good use of coping tactics to reach the lowest number of omissions and highest terminological accuracy. Their number of omissions however increased when using CASSIS compared to using the spreadsheet glossary. The overall terminology accuracy score was higher with CASSIS than without it. This interpreter reported having looked at the screen for numbers and whenever they needed help with vocabulary. They did not read along with the transcription. Some vocabulary that they read from the software screen have caused omissions later in the text and there was a naturalization while reading the transcription. Their number accuracy was 29% higher with CASSIS than with notation.

6.5.6. Language interferences

A pattern that emerged from analysing the recordings, is an increased number of interferences from the source language when using CASSIS, in this case interferences from English into German. Interpreter 3 naturalised "*pandemics"* at 04:20 into the non-existent word *"Pandemik"* but then quickly corrected it to the correct German word "*Pandemie*". Participant 5 naturalised "*vaccine"* into the similar sounding German word "*Vakzin*" which is rarely used, mostly in medical jargon. This participant then self-corrected it to the commonly used German term "*Impfstoff*". Participant 6 has naturalised the term "*state-of-the-art*" into their German interpretation as "*state-of-the-art*" and then continued without any correction. All these participants have reported to having looked at the CASSIS transcript, however without eye tracking data it is unclear if they were looking at the transcript while the above interferences happened. It is unclear if these interferences were caused by CASSIS or if they were just coincidentally happened during the five minutes when the participants were using CASSIS. Only 1 similar interference was observed when interpreters were in the control group, compared to 7 when they were using CASSIS.

7. Discussion of experiment findings

7.1. Study aims

This experiment set out to explore how an automatised glossary and an automatised transcription can affect the simultaneous interpretation product, in particular its terminological accuracy and the accuracy in interpreting numbers. The CASSIS automatised glossary and transcription was used for this investigation. Six simultaneous interpretations by advanced MA students or MA graduates were recorded, interpreting the same ten-minute speech from English into German. They changed between a bilingual word list- type glossary and the CASSIS tool at the midpoint of the speech.

A quantitative analysis of terminological accuracy was carried out based on the accuracy system of Prandi (2019). Terminological accuracy of enumerations in the speech were also investigated since these are expected to cause a high mental load during SI. This scoring system was adapted to analyse the accuracy in interpreting numbers, and data was collected about the types of errors in number accuracy adapted from Desmet, Vandierendonck, and Defrancq, (2019). Non-textual factors were surveyed, as well: preparation, interpreting experience, language combination, and mental fatigue.

The interpreting process was also studied. A hypothesis based on the literature was that using this CAI tool would influence the interpreting process. The questions of this paper regarding the interpreting process was, how would using an automatised glossary tool change the coping tactics that interpreters use when interpreting a particularly challenging passage in the source text? How do preparation and background knowledge, previous interpreting experience, and working language combination affect interpreting with the CAI tool? Therefore, interpretations were recorded and analysed, and the usage of coping tactics adapted from Gile (1995/2009) summarised. Retrospective interviews were conducted immediately after the experiment with each interpreter reflecting on the interpreting process and their experience with the automatic glossary tool.

7.2. Design limitations

The scope of product analysis in this study is consciously limited to terminological analysis, numbers, and applied interpreter coping tactics. One of the main aims of the tested CAI software is to help use the correct terminology. Its automatic glossary matches can be monitored, data can be gained from them. Since the software's features are explicitly aimed at helping interpreters achieve higher terminological precision, this angle of analysis is one of the focuses of this paper.

Yet performance differences within the study could still be the effect of lack of preparation by participants. Answers regarding how well they are prepared could also be biased towards what they think the study author wants to hear. Therefore, even if they feel prepared according to the questionnaire results, they might have adjusted their answers sub-consciously resulting in discrepancies with their actual level of preparation. For this reason, preparation was self-assessed post interpretation as well. If a participant has decreased their self-assessment, that could indicate they felt they lacked some preparation or background knowledge. If a participant increased their self-assessment after the interpreting, that could indicate they felt they had prepared well.

The study is relying on participant's memory to collect data on how often they were looking at the glossary and at CASSIS. This aspect could be evaluated more accurately using eye-tracking devices.

Analysis of the interpreting process with retrospective interviews is perhaps the biggest methodological limitation of the present study. Subconscious processes and automatisms cannot be articulated in retrospect, and even conscious thought processes are subject to being unretrievable from long-term memory after interpreting (Shamy and Ricoy, 2017). The retrospective interviews in this paper are mainly aimed at *coping tactics,* and tactics are implicitly applied consciously, therefore participants should be able to articulate them. On the other hand, coping tactics can be automatised (Gile, 2015). Therefore, loss of retrospective information about these tactics could happen due to sub-conscious thought processes. To balance out this effect, recordings of the interpretations were analysed to gain more qualitative data about the interpreting process.

The act of recalling thought processes can alter information about them. "*In the act of verbalization mental structures undergo processes such as abstraction, generalization, elaboration and rationalization*" (Shamy and Ricoy, 2017: 54). Furthermore, information recollection may also be skewed by interpersonal dynamics, such as interviewees feeling as though they are being "interrogated" or feeling that their interpreting performance is being scrutinised. (ibid.: 56) Therefore, study participants were given a disclaimer in the beginning of the interviews that their performance is not under evaluation, and that the aim of the study is rather to evaluate the CAI tool with automatic glossary and transcription.

7.3. Quantitative findings

7.3.1. Summary

The data collected about terminological accuracy shows a difference between the two glossary methods. Having access to the automatic glossary and transcription yielded in average 7% higher terminological accuracy scores compared to a bilingual word list glossary. However, 2 out of 6 interpreters have lost terminological accuracy when they had access to CASSIS by 4,8% and 5,5% respectively. 4 out of 6 interpreters have gained accuracy with the tested CAI tool by 20,07%, 15,17%, 11,19% and 6,32% respectively.

Data about the terminological accuracy of enumerations shows an average increase of 9,82% when interpreters had access to the CASSIS glossary tool. Two out of six interpreters were less accurate in this metric when they had access to CASSIS by -43,21% and -1,43% respectively. Four out of 6 study participants interpreted enumeration terminology more accurately with CASSIS, by 27,5%, 20%, 55,71%, and 0,36% respectively.

All interpreters increased their accuracy with numbers when they had access to the automatic transcription, by 38,24%, 46,32%, 7,35%, 16,91%, 2,94%, and 29,41% respectively which means an average increase of 23,53%. Approximations, generalisations, and digit mix-ups were eliminated with the use of automatic transcription. Omissions were reduced by 28% and numbers in wrong orders of magnitude were reduced by 66%.

When cross-examining the above findings about accuracy with non-textual factors, possible correlations were found with language combination, simultaneous interpreting experience, and preparation. No correlation was found between accuracy

and the mental fatigue reported by participants, the frequency of them looking at the CASSIS screen, or the time spent in preparation for the experiment.

7.3.2. Accuracy of interpreted terminology

The data suggests an automatic glossary system increasing interpreter's terminological accuracy on average, although this should be verified by further experiments due to the small scale of the experiment. The data also suggests that using an automatic transcription and automatic glossary query may in some cases decrease terminological accuracy. Several possible correlating factors to such a decrease were identified by this paper: mistakes in the ASR transcription, the graphical interface of the CAI tool used in this experiment, interpreter experience and the way the interpreters utilise the automatic glossary query. A more representative study size would be needed to verify how much these factors contribute to a worse terminological accuracy when using such a CAI tool.

First, matches from the term base are only shown if the ASR system correctly transcribes the source text. With an error percentage of over 7%, this can be a problem and there were indeed cases in the experiment where a key term of a sentence was faultily transcribed and that lead to either the interpreter misunderstanding the message or the interpretation being misunderstandable for listeners. Errors in the transcription however do not explain the problem entirely. The term "*European tech sovereignty*", which was faultily transcribed, was also misunderstood by 2 out of 3 interpreters who did not have access to the transcription during this sentence. In this case, the accent is likely to have caused the misunderstandings. Yet, 2 out of 6 interpreters had a worse terminology accuracy with the automatic glossary than without it.

Secondly, following or frequently looking at the ASR transcription was a strategy that did not seem to yield good results. Mistakes in the ASR system were reported to be distracting. One participant had to divert their mental capacities into trying to find out what an erroneously transcribed word meant, however abandoned their attempt due to the time constraint during SI. Two translators were distracted by the fact that the ASR transcription kept correcting itself on the go.

Thirdly, the 2 interpreters who reported having looked for terminology primarily in the automatic transcription, had worse accuracy percentages than when they had no access to the CAI tool at all. This seems to correlate with the fact that source language terminology was highlighted in bright yellow in the automatic transcription, but the matches from the term base were not highlighted. The strategy that seemed to work better was not to follow along with the transcription. Instead, if interpreters needed the help of the glossary, they looked on the right side of the screen where matches from the term base were shown in the source as well as the target language. 2 out of 6 interpreters used this glossary strategy from the get-go, another 2 switched to this strategy after they were initially distracted by mistakes or corrections in the ASR transcription. All 4 interpreters who reported having looked mainly at the suggested matches from the term base when they needed terminology help, have increased their terminological accuracy when they had access to the CAI tool. The effectiveness of this strategy depends on the uploaded term base.

The fourth possible factor is that the interpreters who recorded a lower terminological accuracy when they had access to CASSIS were both already practicing interpreters. A bigger sample is needed to verify whether there is a causal connection between these two factors. It is possible that they found it harder to adjust to using an automatic glossary because they were already used to their own method of glossary query. It is also possible that because of their professional experience, they could rely more on their interpreting coping tactics other than consulting resources in the booth, such as explaining, paraphrasing, reconstructing from context, generalising to yield better results. Meanwhile the in-depth analyses suggest that these coping tactics were somewhat hampered by the availability of CASSIS on screen, thus also reducing the accuracy of those who otherwise would have relied more on those tactics.

Another possible factor is preparation, although no correlation could be determined between the accuracy and the preparation and background knowledge self-assessment after the experiment. However, this study design only measured preparation via a questionnaire and no correlation could be found between accuracy results and the preparation surveyed in the questionnaire. Studies that include more in-depth tests on interpreter’s preparedness on the topic could deliver more accurate results in this regard.

Every interpreter received the same glossary (see Appendix) which was compiled based on the source speech. The glossary entries also equalled the term base entries in the automatic glossary of CASSIS. This means that all the specialised terminology in the source speech was entered into the term base. This in turn is different from the preparation process for an interpreting assignment (Moser-Mercer 1992, Will 2007) and means that the glossary in this experiment was unrealistically complete. Expressions such as "*European tech sovereignty*" were included, which is not a common expression at all. The main reason for this was to control a variable and make sure all interpreters have the same glossary. Another reason was to gain as much data as possible on how CASSIS affects interpretation. That is why all specialised terminology from the source text was included in the glossaries. In a real-world scenario this would not be the case. An important term missing from the term base is a possibility for which interpreters should be prepared if they ever use such a CAI tool for an assignment.

7.3.3. Accuracy of interpreted numbers

The data suggest that the accuracy of interpreted numbers improves if an automatic transcription can be used. All interpreters have improved their number accuracy score when they had access to the CASSIS screen, in average by 23,53%. Every type of mistaken interpretation of numbers was reduced with the use of the automatic transcription.

The fact that approximations, generalisations, and digit mix-ups were eliminated completely with the use of automatic transcription points to better accuracy when using such a system. Omissions of numbers were still observed (although reduced by 28%), and this could mean that the system can be improved. As suggested by the interviews, and already implemented in the InterpretBank number recognition feature, numbers could be written out separately from the transcription, thus making them easier to locate visually and reducing the cognitive load on the interpreters. At the same time, it was found during the in-depth analyses, that omissions in a sentence cannot always be pinpointed to something that went wrong during the same sentence. They can occur due to problems such as too long ear-to-voice span from a problem and subsequent delay in a previously interpreted sentence.

7.3.4. Comparing results with previous studies

The measured terminological accuracy in this study is similar to the results found by Prandi (2019). Her data suggested that a CAI tool which automatically presents simultaneous interpreters with suggested target language terminology improved accuracy by 4 out of 6 interpreters. The remaining 2 out of 6 in her experiment achieved a slightly higher terminological accuracy using a glossary in Word or Excel then with an automatised query in InterpretBank. She thus concluded that the results vary too much from subject to subject to draw conclusions about terminological accuracy.

This paper yielded similar results to Prandi's experiment, in that 4 out of 6 interpreters have improved their terminological accuracy using the tested CAI tool, and 2 out of 6 had a better accuracy in the control group when they had only a spreadsheet glossary.

The participant group and the source text are key differences between the present experiment and the cited study by Prandi. The participant group in the present experiment was more diverse with four MA students and 2 MA graduates with less than 5 years of professional experience. The interpreters in the cited study were all MA students. A 10-minute speech from the EU Parliament was interpreted in the present paper, whereas in the cited paper it was a source text purpose-written for the experiment. These differences make it harder to compare results, but the similar improvement ratio can still be noteworthy.

The study about interpreting numbers with the help of an ASR system by Desmet, Vandierendonck, and Defrancq (2019) produced an increase of number accuracy of around 30%. The present study produced an average number accuracy increase of 23%. This difference could have been caused by the fact that the cited study used a mock-up version of an ASR system, that displayed numbers immediately as they were uttered within the text. An automatic transcription of a live text inherently has a delay, which also could have contributed to the smaller increase in accuracy of numbers than in the cited study. Numbers in the above cited study were displayed on their own and always written out with digits. In the Google Cloud text-to-speech system, numbers were sometimes written out with digits, but sometimes spelled out in the transcription. This makes it harder to anticipate if interpreters will have to find digits

or words in the transcription. The advantage is that units of measurement will also be in the transcription context automatically. The ASR in the present study transcribed all numbers and units correctly.

7.4. Qualitative findings

7.4.1. Summary

The recorded simultaneous interpretations suggest that while having access to an automatic transcription and automatic term base matches reduces the frequency of omitting and generalising information due to high cognitive load, the product becomes more segmented, with more interferences from the source language.

Findings from the interviews suggest that improvements in the visual layout of the investigated CAI tool, in particular highlighting the automatically suggested matches from the term base, highlighting numbers, and writing them out with digits could lead to better usability and thus more trust by the users. The interviews also suggest that mistakes in the automatic transcription make users lose their concentration and thus lead to mistakes and omissions.

Analyses of individual recordings show that using this automatic glossary may have very different effects on the product, as well as the process depending on the user and their preferred way of interpreting.

7.4.2. Effect on coping tactics

The data suggests that using an automatic transcription and glossary during SI can change the interpreting process. The coping tactics used during high mental loads, adapted from Gile (2009), were analysed in the recorded interpretations.

The analysis showed a decrease in the number of omissions from 46 to 31 across all 6 recorded interpretations in total, which amounts to 32% average decrease in omissions. A decrease in generalisations from 33 to 25 was measured meaning a 24% average decrease. This suggests less information loss during SI when using the automatic transcription and glossary.

The analysis also showed that with the availability of CASSIS tool, the number of language interferences from the source language went up from 0 to 10, with 8 naturalisations and 2 transcoded structures recorded in total across 6 interpretations.

Based on the post-experiment interviews, there is a possibility that these were caused by the source language transcription present on screen, which could be further verified by future studies using eye-tracking data, as well as further studies into the priming effect of accessing an ASR transcript while interpreting.

The frequency of segmentation as a tactic also increased from 6 to 12 in total across 6 interpretations when the CASSIS tool was on screen. The following interpreter coping tactics were also analysed, but no significant usage differences were found on average between the CAI tool and the spreadsheet glossary: delaying the response, reconstructing a segment from context, changing the ear-voice span, changing the order of elements in an enumeration, paraphrasing, explaining.

However, analysing the recording one by one for individual interpreter preferences paints a different picture. Using CASSIS has in some cases meant a preference for segmentation and consulting resources in the booth, while a decrease in other coping tactics such as paraphrasing, explaining, reconstructing a segment from context. Looking up terminology and numbers from the CAI tool caused more audible hesitations in the interpretation than in the control group.

This data suggests that using an automatic glossary and transcription CAI tool can alter the SI process. Interpretations with the CASSIS tool tended to have less information loss through a decrease in omissions and generalisations, but the changes in coping tactics tendencies also meant a more segmented, less fluent product with more language interferences such as naturalising words, using a foreign word order, wrong prepositions. Since this study used a small sample of 60 minutes interpretation in total and data was gained only from retrospective interviews and from analysis of the recorded interpretations, this data should be further verified by bigger studies and using eye-tracking data.

7.4.3. In-depth analyses of interpretations of segments with an increased load on mental capacities

Interpretations of some particularly challenging segments were analysed in-depth to further explore the process of SI with the CAI tool in question. According to the Effort Model (Gile 1995/2009), specialised terminology and numbers can create a

high mental load during SI, therefore interpretations of two passages were analysed in-depth.

The in-depth analyses of two sentences containing 3 specialised terms and one year at minute 07:20 in the source text suggests that looking up terminology during SI from the automatic term base matches can result in a more complete translation than looking up terminology from a spreadsheet-style glossary. A term search from the glossary took so much time that the interpreter ended up omitting the rest of the two sentences. Another interpreter who did not have the automatic glossary during this segment, substituted the key term with an adverb that made both sentences potentially misunderstandable for listeners. Interpreters who had CASSIS on screen and reported to having used it to help with terminology had more complete sentences with less information lost. There was an audible pause at each of the 3 term suggestions shown on the CASSIS screen in the SI product of interpreter 2. This interpreter incorporated all three term suggestions into three short sentences, however with some lack of precision due to the use of wrong auxiliary verbs. The result was very segmented, but almost complete interpretation. Interpreter 4 used a coping tactic to delay the interpretation and incorporated only 2 out of 3 suggested terms from the CASSIS screen. This in-depth analysis suggests that using CASSIS can indeed help achieve a more complete translation with segments that have a lot of specialised terminology, however, there were some audible “side effects” in the product. Hesitations and pauses were observed during the analysed SI products along with some imprecision in the auxiliary verb choice and language interferences such as an irregular word order even by the participant who was interpreting into their A language.

An in-depth qualitative analysis of two short sentences containing 3 numbers at 04:33 in the source speech is an example for the accuracy differences shown in the quantitative data. The 3 interpreters who had CASSIS on screen during this segment interpreted the correct number in total 7 times and missed the number in total 2 times. The ones who did not have CASSIS interpreted numbers 5 times correctly, one number was generalised, one number was interpreted with its digits mixed up and two numbers were missed. Two people had an apparent contradiction in their SI product, which roughly can be translated back to “*42 European countries on 4 different continents*”. The interpreter who did not have CASSIS on screen managed to self-correct this contradiction by starting a new sentence and the interpreter with CASSIS

on screen continued without correction. This seems to suggest a better accuracy in interpreting numbers with the live transcription, while further studies are needed to verify the seeming decline in self-reflection capacities when having CASSIS on screen.

7.4.4. Automatic speech recognition, the Achilles' heel

Usability of the tested automatic transcription and automatic term base matches depends by design largely on the automatic speech recognition (ASR) engine that is running underneath the software.

The ASR in the tested software was the Google Cloud speech-to-text engine. It did not perform as recommended. During the test, input audio was ensured to have a high quality with an external sound card and aux cable to input sound – as opposed to a microphone which would have had lower input quality. The Google ASR system that is running in the CASSIS software tested in the present paper achieved an average word error rate of 7,4%, which is higher than the 5% recommended value by Fantinuoli (2019), which can be achieved by for example Dragon NaturallySpeaking ASR engine in English. However, the speaker in the present experiment was not a native English speaker which likely worsened the performance of the ASR.

Analyses of two interpretation segments were conducted in which the automatic transcription was erroneous. In both analysed segments, the speaker pronounced the words with a somewhat heavier accent and was using the acronyms EU and SURE. The recorded interviews suggest that in these cases, the accuracy of interpreters who had CASSIS on screen was lower than those who did not have it. In all but one of the analysed cases, interpreters with an access to the automatic transcription missed the terms that the transcription did not pick up and failed to use any of the other coping tactics that would have been available. The other interpreters who did not have CASSIS on screen used the following coping tactics in the same segments: generalising, reconstructing from context, paraphrasing, and explaining. This suggests that interpreters using CASSIS in this experiment had shifted their coping strategies to looking at the screen and therefore put less emphasis on other available tactics. These results suggest that choosing the ASR engine with the lowest possible error rate is crucial when designing an automatic glossary or transcription system for interpreters.

8. Conclusions and further research

CAI tools that use automatic speech recognition (ASR) to suggest term translations, and to note down numbers, are a recent development in interpretation (Fantinuoli 2017). Data from this study about terminological accuracy and number interpretation accuracy was similar to the key findings of two previous studies about SI performance with CAI tools based on ASR (Prandi 2019, Desmet, Vandierendonck, and Defrancq 2019). Due to some methodological differences, the present study cannot be considered a confirmation of the previous two, but the results are similar. Based on the data, CAI tools can improve accuracy of numbers during SI, and in some cases can improve terminological accuracy.

Analysis of the problem sections in interpretation suggests that there are effective and less effective ways to use the tested CAI tool. While in most cases, interpretation accuracy has improved, in some cases, using the tested CAI tools had a detrimental effect on accuracy. Two out of 6 participants had an overall worse accuracy when they were using the tested CAI tool compared to their results in the control group. Based on the data, the three main factors that correlate with lowered accuracy of these interpreters were identified. Continuously reading along with the automatic transcription, looking for automatically suggested translations in the wrong place on the software screen, or their previous interpreting experience and their way of handling problems with coping tactics. How frequently interpreters look at the screen and where they look for suggestions might influence the interpreting product.

There are also factors beyond the interpreter's control that can influence SI accuracy. Firstly, whenever the automatic transcription contained errors in the key terms of a sentence, accuracy on average has deteriorated, even compared to just using a spreadsheet glossary. Therefore, this data suggests that choosing the best available ASR engine for any such CAI tool is crucial. Secondly, post-interpretation interviews suggest that good graphical design, such as intuitive and easy-to-spot visual cues to numbers in the source text and suggestions from the term base, could make the tested CAI tool more effective. The small study size however means it is not enough to draw conclusions as to what makes the tested CAI tool work best.

Further research is therefore needed to assess how or why automatic glossary matches, and transcriptions can influence accuracy. Using eye tracking data could

more accurately determine how visual attention to specific software features such as the suggested translations, transcription or highlighted words affect the translation product. Using more elaborate ways than what was available in this study to measure mental fatigue, to study the interpreting process and the preparation of interpreters could shed light on how the tested CAI tools affect the interpreting process.

Bibliography

Bowker, L. (2019) ‘Terminology’, *Routledge Encyclopedia of Translation Studies*. 3rd edn. London: Routledge.

Bühler, H. (1986) ‘Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters’, *Multilingua*, 5(4), pp. 231–235.

Cammoun, R. *et al.* (2009) *Simultaneous Interpretation with Text. Is the Text ‘Friend’ or ‘Foe’? Laying Foundations for a Teaching Module*. University of Geneva.

Corpas Pastor, G., Costa, H. and Durán-Muñoz, I. (2014) ‘A Comparative User Evaluation of Terminology Management for Interpreters', In: Drouin, P., Grabar, N., Hamon, T. and Kageura, K. (eds.), *Proceedings of the 4th International Workshop on Computational Terminology*. Dublin, pp. 68–76.

Corpas Pastor, G. and Fern, L.M. (2016) *A survey of interpreters’ needs and practices related to language technology*. FFI2012-38881-MINECO/TI-DT-2016–1. University of Malaga.

Costa, H., Corpas Pastor, G. and Durán-Muñoz, I. (2014) ‘Technology-assisted Interpreting’, *MultiLingual*, 143,25(3), pp. 27–32.

Defrancq, B. and Fantinuoli, C. (2020) ‘Automatic speech recognition in the booth Assessment of system performance, interpreters’ performances and interactions in the context of numbers’, *Target. International Journal of Translation Studies*, 33(1), pp. 73–102. Available at: https://doi.org/10.1075/target.19166.def.

Derwing, T.M., Munro, M.J. and Carbonaro, M. (2000) ‘Does Popular Speech Recognition Software Work with ESL Speech?’, *TESOL Quarterly*, 34(3), pp. 592–603.

Desmet, B., Vandierendonck, M. and Defrancq, B. (2019) ‘Simultaneous interpretation of numbers and the impact of technological support.' In: Fantinuoli, C. (ed.), *Interpreting and technology*. Berlin: Language Science Press, pp. 13–27. Available at: https://doi.org/10.5281/zenodo.1493293.

Díaz-Galaz, S., Padilla, P. and Bajo, M.T. (2015) ‘The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students’, *Interpreting*, 17(1), pp. 1–25.

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014) *Internet, phone, mail, and mixed-mode surveys : the tailored design method*. 4th edn. Hoboken, New Jersey: Wiley.

Fantinuoli, C. (2017) ‘Speech Recognition in the Interpreter Workstation' In: Esteves-Ferreira, J, Macan, J., Mitkov, R, Stefanov, O.-M. (eds.), *Translating and the computer 39 - Proceedings*. London, pp. 25–37.

Fantinuoli, C. (2018) ‘Computer-assisted Interpreting: Challenges and Future Perspectives' In: Corpas Pastor, G. und Durán-Muñoz, I. (eds.),*Trends in E-Tools and Resources for Translators and Interpreters*. Leiden: Brill, pp. 153–174. Available at: https://doi.org/10.1163/9789004351790_009.

Fantinuoli, C. (2019) ‘Interpreting and technology: The upcoming technological turn' In: Fantinuoli, C. (ed.), *Interpreting and technology*. Berlin: Language Science Press, pp. 1–12.

Gile, D. (1995) *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: John Benjamins.

Gile, D. (2009) *Basic Concepts and Models for Interpreter and Translator Training: Revised Edition*. Amsterdam: John Benjamins.

Gile, D. (2017) ‘Testing the Effort Models’ Tightrope Hypothesis in Simultaneous Interpreting - A Contribution’, *Hermes*, 23, pp. 153–172.

Gumul, E. (2019) ‘Evidence of cognitive effort in simultaneous interpreting: Process versus product data’, *Beyond Philology: An International Journal of Linguistics, Literary Studies and English Language Teaching*, 16(4), pp. 11–43.

Hajos, S., Hajos, K. and Klein, P. (2020) *CASSIS - Computer Assisted Interpreter System*. Available at: https://www.alexexpert.hu/cassis/ (Accessed: 19 July 2021).

ISO (2019) ‘ISO 20539:2019 Translation, interpreting and related technology — Vocabulary’. BSI Standards Limited. Available at: https://www.iso.org/standard/68293.html.

Ivanova, A. (2000) ‘The Use of Retrospection in Research on Simultaneous Interpreting' In: Tirkkonen-Condit, Sonja and Jääskeläinen, Riitta (eds.), *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on empirical research*. Amsterdam: John Benjamins, pp. 27–52.

Kurz, I. (1993) ‘Conference Interpretation: Expectations of Different User Groups’. Available at: https://www.openstarts.units.it/handle/10077/4908 (Accessed: 23 December 2021).

Kurz, I. (2001) ‘Conference Interpreting: Quality in the Ears of the User’, *Meta: Translators’ Journal*, 46, pp. 394–409.

Lai, J., Karat, C.-M. and Yankelovich, N. (2008) ‘Conversational speech interfaces and technologies' In: A. Sears & J. A. Jacko (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. New York: Erlbaum, pp. 381–391.

Levis, J. and Suvorov, R. (2012) ‘Automatic Speech Recognition' In: Carol A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. 1st edn. Malden, Massachusetts: Wiley.

Mellinger, C. (2019) ‘Computer-Assisted Interpreting Technologies and Interpreter Cognition: A Product and Process-Oriented Perspective’, *Revista Tradumàtica*, 17, pp. 33–44.

Moser-Mercer, B. (1992) ‘Banking on Terminology Conference Interpreters in the Electronic Age’, *Meta*, 37(3), pp. 507–522.

Moser-Mercer, B. (1997) ‘Beyond Curiosity: Can Interpreting Research Meet the Challenge?' In: Danks, Joseph H. (ed.), *Cognitive processes in translation and interpreting.* Thousand Oaks, Calif.: Sage.

Moser-Mercer, B. (2002) ‘Process models of simultaneous interpretation' In: Pöchhacker, F., Schlesinger, M. (eds.), *The interpreting studies reader*. London, New York: Routledge.

Orlando, M. (2010) ‘Digital pen technology and consecutive interpreting: Another dimension in notetaking training and assessment’, *The Interpreters’ Newsletter*, (15), pp. 71–86.

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

Ortiz, L.E.S. and Cavallo, P. (2018) ‘Computer-Assisted Interpreting Tools (CAI) and options for automation with Automatic Speech Recognition’, *Tradterm*, 32, pp. 9–31.

Pöchhacker, F. (2016) *Introducing Interpreting Studies*. 2nd edn. London: Routledge.

Prandi, B. (2017) ‘Designing a Multimethod Study on the Use of CAI Tools during Simultaneous Interpreting' In: Esteves-Ferreira, J, Macan, J., Mitkov, R, Stefanov, O.-M. (eds.), *Translating and the Computer 39 - Proceedings*. London.

Prandi, B. (2019) ‘An exploratory study on CAI tools in simultaneous interpreting: Theoretical framework and stimulus validation' In: Fantinuoli, C. (ed.), *Interpreting and technology*. Berlin: Language Science Press, pp. 29–59. Available at: https://doi.org/10.5281/zenodo.1493293.

Pym, A. (2011) ‘What technology does to translating’, *Translation & Interpreting*, 3(1), pp. 1–9.

Riccardi, A. (2015) ‘Speech Rate’, *Routledge Encyclopedia of Interpreting Studies*. London: Routledge.

Rodríguez, N. and Schnell, B. (2009) ‘A Look at Terminology Adapted to the Requirements of Interpretation’, *Language Update*. 6th edn, p. 21.

Sandrelli, A. (2015) ‘Computer-Assisted Interpreter Training’, *Routledge Encyclopedia of Interpreting Studies*. London: Routledge.

Seeber, K.G. (2011) ‘Cognitive load in simultaneous interpreting: Existing theories — new models.’, *Interpreting*, 13(2), pp. 176–204.

Seeber, K.G. (2017) ‘Multimodal Processing in Simultaneous Interpreting' In: Schwieter, J. W. and Ferreira, A. (eds.), *The Handbook of Translation and Cognition*. Hoboken: John Wiley & Sons.

Setton, R. (2011) ‘Corpus-Based Interpreting Studies (CIS): Overview and Prospects' In: Kruger, Alec (ed.), *Corpus-Based Translation Studies*. London: Continuum.

Setton, R. and Motta, M. (2007) ‘Syntacrobatics: Quality and reformulation in simultaneous-with-text’, *Interpreting*, 9(2), pp. 199–230.

Shamy, M. and Ricoy, R. (2017) ‘Retrospective protocols: Tapping into the minds of interpreting trainees’, *Translation and Interpreting*, 9, pp. 51–71.

Stewart, C. *et al.* (2018) ‘Automatic Estimation of Simultaneous Interpreter Performance' In: Artzi, Y. and Eisenstein, J. (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, pp. 662–666.

Tripepi Winteringham, S. (2010) ‘The usefulness of ICTs in interpreting practice’, *The Interpreters’ Newsletter*, 15, pp. 87–99.

Viglino, T., Motlicek, P. and Cernak, M. (no date) ‘End-to-End Accented Speech Recognition’, *Interspeech 2019*, pp. 2140–2144. Available at: https://doi.org/10.21437/Interspeech.2019-2122.

Vogler, N., Stewart, C. and Neubig, G. (2019) *Lost in Interpretation: Predicting Untranslated Terminology in Simultaneous Interpretation*. Available at: https://www.researchgate.net/publication/332139073_Lost_in_Interpretation_Predicting_Untranslated_Terminology_in_Simultaneous_Interpretation (Accessed: 5 June 2021).

Wickens, C.D. (1984) ‘Processing resources in attention' In: R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention*. New York: Academic Press, pp. 63–102.

Wickens, C.D. (2002) ‘Multiple resources and performance prediction’, *Theoretical Issues in Ergonomics Science*, 3 (2).

Will, M. (2007) ‘Terminology Work for Simultaneous Interpreters in LSP Conferences: Model and Method' in: Gerzymisch-Arbogast, H. and Budin, G. (eds.), *Proceedings of the Marie Curie Euroconferences. MuTra: LSP Translation Scenarios*. *Multidimensional Translation - MuTra*, Vienna, p. 10. Available at: https://www.euroconferences.info/proceedings/2007_Proceedings/2007_proceedings.html.

Will, M. (2015) ‘Zur Eignung simultanfähiger Terminologiesysteme für das Konferenzdolmetschen’, *trans-kom*, 8(1), pp. 179–201.

Xu, R. (2018) ‘Corpus-based terminological preparation for simultaneous interpreting’, *Interpreting*, 20(1), pp. 29–58.

Appendix

Zusammenfassung

Computergestütztes Dolmetschen (im Englischen: computer-assisted interpreting, CAI) wird in der Zukunft an Stellenwert gewinnen (vgl. Pym, 2011; Corpas Pastor, Costa and Durán-Muñoz, 2014; Corpas Pastor and Fern, 2016; Fantinuoli, 2018, Vogler, Stewart and Neubig, 2019). Unterschiedliche Zwecke des CAI werden in dieser Masterarbeit kurz zusammengefasst.

Der Schwerpunkt sind CAI-Software mit einer automatischen Spracherkennungssystem (im Englischen: automatic speech recognition, ASR) und einer darauf basierenden automatischen Glossarfunktion (vgl. Fantinuoli, 2017). Diese Art von Software kann laut Fachliteratur einerseits zu einer höheren Präzision beim Dolmetschen von Terminologie und Zahlen beitragen (vgl. Fantinuoli, 2017; Desmet, Vandierendonck and Defrancq, 2019; Prandi, 2019), andererseits könnten sie durch eine erhöhte mentale Anstrengung beim Dolmetschen (vgl. Gile, 2007; Seeber, 2017) die Dolmetschung auch negativ beeinflussen (vgl. Stewart et al., 2018).

Diese zwei verschiedenen Ansätze in der Fachliteratur begründen eine Untersuchung des Dolmetsch-Produktes und der kognitiven Prozesse während des Dolmetschens (vgl. Mellinger,2019; Prandi 2019). Prandi (2019) schlägt vor, dass zukünftige Studien über CAI-Software mit automatischer Spracherkennung auch eine automatisierte Glossar-Funktion untersuchen sollten, was das Ziel dieser Masterarbeit ist. In dieser Masterarbeit wird daher die CAI Software CASSIS untersucht (Hajos, Hajos and Klein, 2020).

Die erste Forschungsfrage ist, wie die getestete Software die gedolmetschte Terminologie und Zahlen beeinflusst. Die Methodologie basiert auf der Studienstruktur von Prandi (2019), die zur Untersuchung dieser Art von CAI-Software entworfen wurde. Eine 10-minütige Rede wurde aus dem Englischen ins Deutsche gedolmetscht und die gedolmetschte Fachterminologie wurde analysiert. Davon wurden pro Studienteilnehmer 5 Minuten ohne Software gedolmetscht und 5 Minuten mit der CASSIS Software. Die Präzision wurde nach dem Punktesystem von Prandi (2019) untersucht.

Zweitens wurde der Dolmetschprozess mit CASSIS untersucht, mit Fokus auf die Strategien, die beim Dolmetschen genutzt werden. (vgl. Strolz, 1994; Gile, 2009; Seeber, 2011) Die Dolmetschungen wurden aufgenommen, die verwendeten Dolmetschstrategien sowie die Ursachen der Probleme nach Möglichkeit analysiert und die Analyse mit den Ergebnissen aus einem Fragebogen und Interviews (vgl. Ivanova, 2000) verglichen.

In Punkto Fachterminologie waren die Dolmetschungen mit CASSIS im Durchschnitt 7% pünktlicher als die in der Kontrollgruppe. Jedoch verbesserten nur 4 von 6 Dolmetscher*innen ihre terminologische Pünktlichkeit, um jeweils 20,07%, 15,17%, 11,19% und 6,32%. 2 von 6 Dolmetschungen waren mit CASSIS terminologisch unpünktlicher, um jeweils 4,8% und 5,5%. In Bezug auf gedolmetschte Nummer hat das untersuchte System zu einem Anstieg in Pünktlichkeit von durchschnittlich 23,53% geführt. Die Nummer waren in jeder einzelnen Dolmetschung mit CASSIS pünktlicher als in der Kontrollgruppe.

Die Analyse der Dolmetsch-Strategien zeigte, dass ein automatisches Transkript zu einem Rückgang in der Anzahl an Auslassungen und Generalisierungen beitragen kann, jedoch werden die Dolmetschungen dadurch segmentierter und weisen mehr Interferenz aus der Ausgangssprache auf. Laut den nach der Dolmetschung durchgeführten Interviews können Fehler in der automatischen Transkription zu einem Verlust von Vertrauen in das CAI-Tool bei den Dolmetschern führen. Eine verbesserte graphische Oberfläche der Software in Bezug auf die automatische Darstellung der Wortpaare aus dem Glossar, und die Darstellung der Nummer könnte zu einer weiteren Verbesserung der Ergebnisse mit der Software führen.

Die Ergebnisse dieser Studie sind ähnlich zu vorigen Studien über CAI-Software mit automatisierter Spracherkennung (vgl. Prandi 2019, Desmet, Vandierendonck, and Defrancq 2019). Jedoch können aufgrund der Limitierungen dieser Studie nur sehr vorsichtige Schlüsse gezogen werden. Die Teilnehmer*innenzahl ist nicht repräsentativ, und die Analyse basierte auf aufgenommene Dolmetschungen, Interviews und Fragebögen. Mit zusätzlicher Blickerfassungstechnologie und einer größeren Studienzahl könnten weitere Erkenntnisse gewonnen werden.

Questionnaire

1. Gender
    a. Male
    b. Female
    c. Other
2. Age
    a. 20-24
    b. 25-29
    c. 30-39
    d. 40-49
    e. 50 or above
3. Interpreting experience
    a. Student
    b. Novice professional (4 or less years of experience)
    c. Experienced professional (5 or more years of experience)
4. I have studied the following:
    a. Simultaneous interpretation
    b. Consecutive interpretation
    c. Both
    d. In an MA programme
    e. In a BA programme
    f. In another course
5. My working languages are:

_____

6. The type of tools I usually use to prepare for an interpreting assignment are:

_____

7. The type of tools I usually use during interpreting are:

_____

8. My opinion about Gile's four efforts of interpreting (Listening and Analysis, Production, Memory, Coordination) in one sentence:

_____

9. In my experience, coordination of these efforts during SI can be improved by:

_____

10. My preferred method(s) to prepare for an interpreting assignment:

_____

11. I would rate my general background knowledge of politics:
    - 1 2 3 4 5 6 7 8 9 10
12. I would rate my general background knowledge of healthcare:
    - 1 2 3 4 5 6 7 8 9 10

13. I would rate my general background knowledge of technology:
    - 1 2 3 4 5 6 7 8 9 10
14. I would rate my general background knowledge of economics:
    - 1 2 3 4 5 6 7 8 9 10
15. This is how much time I spent preparing for this interpretation:
    - 0-30 min
    - 30 min-1 hour
    - 1-2 hours
    - More than 2 hours:____
16. I feel at this moment:
    - Cognitively exhausted -- -- at my best cognitive capacity
    - 1 2 3 4 5 6 7 8 9 10

After interpretation:

1. I feel at this moment:
    - Mentally exhausted -- -- -- at my best cognitive capacity
    - 1 2 3 4 5 6 7 8 9 10
2. I have heard (yes/no) or interpreted (yes/no) this speech before.
3. I found the source speech
    - easy -- -- -- -- -- -- -- -- difficult
    - 1 2 3 4 5 6 7 8 9 10
4. I would rate my performance in politics:
    - 1 2 3 4 5 6 7 8 9 10
5. I would rate my performance in healthcare:
    - 1 2 3 4 5 6 7 8 9 10
6. I would rate my performance in technology:
    - 1 2 3 4 5 6 7 8 9 10
7. I would rate my performance in economics:
    - 1 2 3 4 5 6 7 8 9 10
8. How often was I looking at CASSIS
    - Only once -- -- sometimes -- often -- -- all the while
    - 1 2 3 4 5 6 7 8 9 10
9. I found interpreting with CASSIS
    - Bad -- -- -- neutral -- -- -- -- good
    - 1 2 3 4 5 6 7 8 9 10
10. I found the UI (user interface) of CASSIS
    - Hard to understand -- neutral -- -- -- easy to understand
    - 1 2 3 4 5 6 7 8 9 10
11. I would rate the usefulness of the automatic transcription in CASSIS
    - 1 2 3 4 5 6 7 8 9 10
12. I would rate the usefulness of the glossary function in CASSIS

- 1 2 3 4 5 6 7 8 9 10

13. CASSIS helped me if I was lagging behind
    - Yes
    - No
    - Partially
14. What I liked about CASSIS

_____

15. What I dislike about CASSIS

_____

16. What I liked about interpreting with a paper glossary

_____

17. What I dislike about interpreting with a paper glossary

_____

18. I would add the following function(s) to CASSIS:

_____

19. Would I use CASSIS in the future?
    - Yes, and here is why:
    - No, and here is why:

Source speech transcription

Honourable members! **

A year is long in/ at the time of pandemics. When I stood here in front of you twelve months ago, I did not know when or even if we could have a safe and effective vaccine against covid-19. But today, * against all critics * Europe is among the world leaders. More than 70% of our adult population is fully vaccinated. We were the only ones to share half of our vaccine production with the rest of the world. We delivered more than seven hundred million doses of vaccines to the Europeans. We delivered, ((5s)) indeed we delivered also more than seven hundred million to the rest of the world, to more than one hundred thirty countries. And we are the only region in the world to achieve that. Honourable members, a pandemic is a marathon it's not a sprint. We followed the science, we delivered to Europe, we delivered to the rest of the world we did it the right way because we did the European way and I think it worked. ((5s))

But while we have every reason to be confident, we have no reason to be complacent. Our first and most urgent priority is to speed up global vaccinations. With less than one percent of global doses administered in low-income countries, the scale of injustice and the level of urgency * is obvious. This is one of the great geopolitical issues in our time. Team Europe * is investing one billion Euro * to ramp up mRNA production capacity with Africa. We have already committed to share two hundred fifty million doses of vaccine. I can announce today that the Commission will add a new donation of another two hundred million doses until the middle of next year. This is an investment in solidarity, and it is an investment also in global health.

The second priority is to continue our efforts here in Europe. We see worrisome divergences between member states, what vaccination rates are concerned. So, we need to keep up the momentum. And Europe is ready! We have one point eight billion additional doses secured. This is enough for us and our neighbourhood * when booster shots are needed. So, let's do everything possible * that this does not turn into a pandemic of the unvaccinated.

And my third point, the final priority is to strengthen our pandemic preparedness. Last year I said it was time to build a European Health Union. * Today we are delivering. * With our proposal, we get the HERA authority up and running. This will be a huge asset to deal with future health threats earlier and better. We have the innovation and

scientific capacity, we have the private sector knowledge, we have competent national authorities, and now we have to bring all of that together including massive * funding. So, I am proposing a new health preparedness and resilience mission * for the whole of the European Union, and it should be backed up by a Team Europe investment of fifty billion by twenty twenty-seven. To make sure that no virus will ever turn a local epidemic again in a global pandemic. There is no better return on investment then that. ((2s))

Honourable members, the work on the European Health Union is a big step forward * and I want to thank this house for your support. We have shown that when we act together, we can act fast. Take the EU Digital Certificate: Today, more than four hundred million certificates have been generated across Europe. Forty-two countries * in four different continents are plugged in. We proposed it in March you pushed hard you helped enormously three months later, three months later only it was up and running. So, while are the rest of the world was talking about Europe * just * did it. ((5s))

And we did a lot of things right. We moved fast to create SURE. This supported over thirty-one billion/ milli/ million workers and two point five million companies across Europe. We learnt the lessons from the past when we were too divided and too delayed. And the difference is stark. Last time it took us eight * years for the Eurozone GDP to get back to pre-crisis level. This time * we expect nineteen countries to be at pre-pandemic levels this year with the rest following next. Growth in the Euro area outpaced both the US and China in the last quarter. But this is only the beginning * and the lessons from the financial crisis could serve as a cautionary tale. At that time Europe declared victory too soon and we paid the price for that. And we will not repeat that mistake. The good news is that with Next-Generation EU we will both invest in short-term recovery and in the long-term prosperity. We will address structural issues in our economy from labour market reforms in Spain to pension reforms in Slovenia or to tax reforms in Austria for example. In an unprecedented manner * we will invest in five G and fibre. But equally important is the investment in digital skills. This is such an important topic. This task needs leader’s attention, and it needs a structure dialogue at top level. We really have to emphasize that and put a strong focus on it. Our response provides a clear direction to market and investors alike. But as we look ahead, * we also need to reflect on how the crisis has affected the shape of our

economy. From increased debts to uneven impact on different sectors or new ways of working. And to do that, * the Commission will relaunch the discussion on the Economic Governance Review in the coming weeks. The aim is to build a consensus on the way forward well ahead of twenty twenty-three.

Honourable members! We will soon celebrate thirty years of single market. For thirty years it has been the great enabler of progress and prosperity in Europe. At the outset of the pandemic, we had to defend it against the pressures of erosion and fragmentation ** for our recovery. The single market is the driver of good jobs and competitiveness and that is particularly important in the digital single market. ** We have made ambitious proposals in the last year for example to contain the gatekeeper power of the major platforms, to underpin their democratic responsibilities of those platforms, to foster innovation, to channel the power of artificial intelligence. Digital is the make * or break * issue, and member states also share that view. For example, if you look at the digital spending in next generation EU, and you know we have a twenty percent target, * we will even overshoot that twenty percent target. That is good. That reflects the importance and/ of investing in our European tech sovereignty. But allow me to focus on semiconductors for moment. These tiny chips that make everything work from smartphones to electric scooters from trains to entire smart factories. There is no digital without those chips. And while we speak, * whole production lines are already working at reduced sp/ speed, despite growing demand, because of a shortage of semiconductors. And while global demand has exploded, Europe’s share across the entire value chain from the design or manufacturing capacities has shrunk. We depend right now on state-of-the-art chips manufactured by Asia. So, this is not just a matter of our competitiveness. This is also a matter of tech sovereignty.

CASSIS bilingual term list

ENGLISH	**GERMAN**
5G	5G
administer	verabreichen
ambitious proposals	ehrgeizige Vorschläge
artificial intelligence	künstliche Intelligenz
booster shot	Auffrischungsimpfung
build a consensus	zu einem Konsens gelangen
certificate	Zertifikat
Commission	Kommission
competent national authority	zuständige nationale Behörde
competitiveness	Wettbewerbsfähigkeit
demand	Nachfrage
democratic responsibility	demokratische Verantwortung
design	Produktgestaltung
digital single market	digitaler Binnenmarkt
digital skills	digitale Kompetenzen
digital spending	Ausgaben für Digitales
doses of vaccine	Impfdosen, die
Economic Governance Review	Überprüfung der wirtschaftspolitischen Steuerung
erosion	Aushebelung (Markt)
EU Digital Certificate	das digitale Zertifikatssystem der EU
European Health Union	Europäische Gesundheitsunion
Eurozone GDP	das BiP in der Eurozone
fibre	Glasfaser
financial crisis	Finanzkrise
foster innovation	Innovation fördern
fragmentation	Fragmentierung
fully vaccinated	vollständig geimpft
funding	Finanzierung
gatekeeper power	Macht über den Markt
geopolitical issue	geopolitisches Thema

global health	weltweite Gesundheit
global pandemic	globale Pandemie
health preparedness and resilience mission	Gesundheitliche Bereitschafts- und Krisensicherheits-mission
health threat	Gesundheitsbedrohung
HERA authority	HERA-Behörde
increased debt	Anstieg der Schulden
labour market	Arbeitsmarkt
local epidemic	lokale Epidemie
long-term prosperity	langfristige Wohlstand
low-income country	einkommensschwaches Land
make-or-break issue	entscheidet über Erfolg oder Scheitern
manufacturing capacity	Fertigungskapazität
member state	Mitgliedstaat
mRNA	mRNA
new ways of working	neue Arbeitsweisen
next-generation EU	Next-Generation EU
our neighbourhood	unsere Nachbarschaft
outpace	überholen
pandemic of the unvaccinated	Pandemie der Ungeimpften
pandemic preparedness	Pandemievorsorge
pandemics	Pandemie, die
pension reform	Rentenreform
plugged in	verbunden
pre-crisis level	Vorkrisenniveau
pre-pandemic level	das Niveau vor der Pandemie
private sector knowledge	Wissen des Privatsektors
production capacity	Produktionskapazität
production line	Fließband
proposal	hier: Vorschlag
propose	vorschlagen
quarter	Quartal

return on investment	eine Geldanlage ist rentabel
scientific capacity	wissenschaftliche Kapazität
sectors	Wirtschaftszweige
semiconductor	Halbleiter
shortage	Knappheit
short-term recovery	kurzfristige Erholung
single market	Binnenmarkt
smart factory	intelligente Fabrik
structural dialogue	strukturierter Dialog
SURE	das SURE-Instrument
tax reform	Steuerreform
Team Europe	Team Europa
tech sovereignty	technologische Souveränität
top level	höchste Ebene
vaccination rate	Impfquote
vaccine	Impfstoff, der
value chain	Wertschöpfungskette