



Hacettepe University Graduate School of Social Sciences

Department of Translation and Interpretation

**AUTOMATIC SPEECH RECOGNITION IN
CONSECUTIVE INTERPRETER WORKSTATION:
COMPUTER-AIDED INTERPRETING TOOL
‘SIGHT-TERP’**

Cihan ÜNLÜ

Master’s Thesis

Ankara, 2023

AUTOMATIC SPEECH RECOGNITION IN CONSECUTIVE INTERPRETER
WORKSTATION: COMPUTER-AIDED INTERPRETING TOOL ‘SIGHT-TERP’

Cihan ÜNLÜ

Hacettepe University Graduate School of Social Sciences
Department of Translation and Interpretation

Master’s Thesis

Ankara, 2023

KABUL VE ONAY

Cihan ÜNLÜ tarafından hazırlanan “Automatic Speech Recognition in Consecutive Interpreter Workstation: Computer Aided Interpreting Tool ‘Sight-Terp’” (Otomatik Konuşma Tanıma Sistemlerinin Ardıl Çeviride Kullanılması: Sight-Terp) başlıklı bu çalışma, 15.06.2023 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Dr. Öğr. Üyesi Alper KUMCU (Başkan)

Prof. Dr. Aymil DOĞAN (Danışman)

Doç. Dr. Gökçen HASTÜRKOĞLU (Üye)

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Uğur ÖMÜRGÖNÜLŞEN

Enstitü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan “**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**” kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- ☐ Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir. ⁽²⁾
- ☐ Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

21/06/2023

Cihan ÜNLÜ

“*Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge*”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez **danışmanının** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez **danışmanının** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulunun** gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, **tezin yapıldığı kurum** tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, **ilgili kurum ve kuruluşun önerisi** ile **enstitü** veya **fakültenin** uygun görüşü üzerine **üniversite yönetim kurulu** tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
- Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez **danışmanının** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** tarafından karar verilir.

ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, **Prof. Dr. Aymil DOĐAN** danıřmanlığında tarafımdan retilildiđini ve Hacettepe niversitesi Sosyal Bilimler Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.

Cihan NL

ACKNOWLEDGEMENTS

My first sincere thanks go to my advisor Prof. Dr. Aymil Doğan. I have nothing but admiration for her wisdom, attention and efforts for her students. I would also like to thank the participants of this study who generously gave their time and energy to take part in this research. Their willingness to share their experiences and insights has been invaluable and deeply appreciated. I would like to express my deepest gratitude to Assoc. Prof. Dr. Didem Tuna and Asst. Prof. Dr. Javid Aliyev for their invaluable academic guidance, understanding, and sincere help throughout both my undergraduate and graduate education. I would also like to extend my thanks to my friend Ebru Kürkcü for her unwavering moral support. I am grateful to Sebahat Gören, Dr. Pınar Uysal Cantürk, Aslı Yolcu and Büşra Ceren Tangül for her kind help in the statistical assessment and evaluation. Lastly, I would like to thank Prof. Didem Tuna and Prof. Alev Bulut for their invaluable expert opinions on the methodology used in this work.

I owe a debt of gratitude to my family, colleagues at Istanbul Yeni Yüzyıl University and friends who provided me with much-needed encouragement, motivation, and support during this challenging time. Their faith in me has been a constant source of motivation and inspiration.

ÖZET

ÜNLÜ, Cihan. *Otomatik Konuşma Tanıma Sistemlerinin Ardıl Çeviride Kullanılması: Sight-Terp, Yüksek Lisans Tezi*, Ankara, 2023.

Bu deneysel çalışma, bilgisayar destekli sözlü çeviri (BDS) aracı olan "Sight-Terp" kullanımının ardıl çeviri sürecine etkisini araştırmaktadır. Bu çalışmanın yazarı tarafından tasarlanan ve geliştirilen Sight-Terp, dijital not defteri, otomatik konuşma tanıma (OKT), gerçek zamanlı konuşma çevirisi, adlandırılmış varlık tanıma ve vurgulama ve otomatik segmentasyon işlevlerine sahiptir. Çalışma, katılımcıların performanslarını iki koşulda (Sight-Terp'li ve Sight-Terp'siz) test etmek ve performanslarını doğruluk ve akıcılık kriterlerine göre analiz etmek için grup içi tekrarlı ölçümler tasarımı kullanmıştır. İki farklı koşuldaki doğruluk oranları arasındaki farkı analiz etmek için doğruluk değişkeni, anlamsal olarak eşdeğer bir şekilde aktarılan anlam birimlerinin sayısının ortalaması ile ölçülmüştür (Seleskovitch, 1989). Akıcılık ise, her bir performans için yanlış başlangıçlar, dolgu duraksamaların sıklığı, sessiz duraksamalar, tüm sözcük tekrarları, bozuk sözcükler ve tamamlanmamış tümceler gibi akıcısızlık göstergelerinin toplam sayısı hesaplanarak ölçülmüştür. Ek olarak, katılımcıların araç kullanımına ilişkin algılarını analiz etmek için deney sonrası anket uygulanmıştır. Elde edilen bulgular, OKT ile entegre edilmiş BDS aracı Sight-Terp'ten yararlanmanın katılımcıların çevirilerinin doğruluğunda bir artışa yol açtığını göstermektedir. Ancak Sight-Terp kullandıklarında katılımcılarda daha fazla akıcısızlık belirteçleri meydana gelmiş ve çeviri için harcadıkları süre görece uzamıştır. Kullanıcılar aracı kullanırken herhangi bir zorluk veya yabancılık hissetmeseler de çalışma sonuçları yazılımın faydasını daha da artırabilecek potansiyel iyileştirme ve değişiklik alanlarını da ortaya koymaktadır. Bu çalışma, OKT teknolojisini sözlü çeviri sürecine dahil etmenin faydalarını ve zorluklarını vurgulayarak sözlü çeviri eğitimi ve pratiğini bilgilendirmeyi amaçlamakta ve sözlü çevirmenler için BDS araçlarının gelecekteki gelişimi için pratik öneriler sunmayı amaçlamaktadır.

Keywords: bilgisayar destekli sözlü çeviri, otomatik konuşma tanıma, sözlü çeviri teknolojileri, ardıl çeviri, not alma, tablet destekli sözlü çeviri

ABSTRACT

ÜNLÜ, Cihan. *Automatic Speech Recognition in Consecutive Interpreter Workstation: Computer-Aided Interpreting Tool ‘Sight-Terp’*, Master’s Thesis, Ankara, 2023.

This experimental study investigates the effect of using an automatic speech recognition (ASR)-enhanced computer-assisted interpreting (CAI) tool “Sight-Terp” on the performances of a group of participants in consecutive interpreting tasks. Sight-Terp, which is designed and developed by the author of this study, provides a digital note-pad, real-time speech translation, named entity recognition and highlighting, and automatic segmentation of a speech. The study employs a within-subjects repeated measures design to test participants' performances in two conditions (with and without Sight-Terp) and analyses their performances based on the criteria of accuracy and fluency. In seeking the significant difference between the accuracy ratios in two different conditions, accuracy was measured by the average of the number of accurately conveyed units of meaning (Seleskovitch, 1989). Fluency, on the other hand, was measured by calculating the total number of occurrences of disfluency markers such as false starts, frequency of filled pauses, filler words, whole-word repetitions, broken words, and incomplete phrases for each performance. Additionally, a follow-up qualitative survey is conducted to obtain participants' comparative responses and perceptions of the tool usage. The analysis and quantitative results of the study indicate that leveraging the ASR-integrated CAI tool Sight-Terp led to an enhancement in the accuracy of the participants' interpretations. However, this also resulted in a higher occurrence of disfluencies and elongated durations of interpretations. While the users experienced little difficulty while using the tool, the study outcomes also suggest potential areas of improvement and modifications that could further enhance the utility of the tool. The study aims to inform interpreting education and practice by highlighting the benefits and challenges of incorporating ASR technology in the interpreting process and offers practical suggestions for the future development of CAI tools for interpreters.

Keywords: computer-assisted interpreting, automatic speech recognition, interpreting technology, consecutive interpreting, note-taking, tablet interpreting

TABLE OF CONTENTS

KABUL VE ONAY	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYAN.....	iii
ACKNOWLEDGEMENTS.....	iv
ÖZET.....	V
ABSTRACT	vi
TABLE OF CONTENTS.....	vii
LIST OF ABBREVIATIONS	xi
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF CHARTS	xv
INTRODUCTION.....	1
CHAPTER 1: SCOPE OF THE STUDY.....	5
1.1. AIM OF THIS STUDY	5
1.2. SIGNIFICANCE OF THIS STUDY	5
1.3. RESEARCH QUESTION(S)	6
1.4. LIMITATIONS	7
1.5. ASSUMPTIONS.....	7
1.6. DEFINITIONS	8
CHAPTER 2: THEORETICAL BACKGROUND.....	9
2.1. INTERPRETING: AN OVERVIEW.....	9
2.1.1. Defining Interpreting.....	10
2.1.2. History of Interpreting.....	11

2.1.3. Interpreting in Modern Times	13
2.1.4. Modes and Settings of Interpreting	15
2.1.4.1. Consecutive Interpreting	18
2.1.4.2. Simultaneous Interpreting	20
2.1.4.3. Sight Interpreting.....	21
2.1.4.4. Whispering (Chuchotage)	22
2.1.4.5. Sign Language Interpreting	23
2.2. EFFORT MODELS IN INTERPRETING.....	23
2.2.1. Effort Models in Consecutive Interpreting.....	25
2.2.2. Effort Models in Human-Machine Interaction	26
2.3. TECHNOLOGY AND INTERPRETING.....	29
2.3.1. The Emergence of Information Technologies in Interpreting.....	29
2.3.1.1. Categorization of Technologies in Interpreting	30
2.3.2. Computer-Assisted Interpreting Tools	37
2.3.2.1. InterpretBank.....	41
2.3.2.2. Kudo Interpreter Assist	44
2.3.2.3. SmarTerp	46
2.3.3. Speech Technologies and Automatic Speech Recognition	48
2.3.3.1. ASR Integration into Translation	53
2.3.3.2. ASR Integration into Interpreting	56
2.3.4. Technology and Consecutive Interpreting	61
2.1.4.1. Sim-Consec	62
3.1.4.1. Tablet Interpreting.....	63
2.4. SIGHT-TERP	65
2.4.1. General Features.....	66
2.4.1.1. Automatic Speech Recognition and Speech Translation	67

2.4.1.2. Automatic Text Segmentation.....	69
2.4.1.3. Named Entity Recognition and Highlighting.....	71
2.4.1.4. Digital Notepad	73
CHAPTER 3: METHODOLOGY	76
3.1. DESIGN OF THE STUDY.....	76
3.2. DATA COLLECTION INSTRUMENTS.....	77
3.2.1. Speeches	78
3.2.2. Questionnaires	81
3.3. PARTICIPANTS.....	82
3.4. PROCEDURE	83
3.4.1. Training	86
3.4.2. Preliminary test	87
3.5. DATA ANALYSIS TECHNIQUES	89
CHAPTER 4: FINDINGS AND DISCUSSION	91
4.1. FINDINGS AND DISCUSSION RELATED TO THE ACCURACY DIFFERENCES	91
4.2. FINDINGS AND DISCUSSION RELATED TO THE FLUENCY DIFFERENCES	93
4.3. POST-EXPERIMENT QUESTIONNAIRE RESULTS.....	96
CONCLUSION AND RECOMMENDATIONS	105
BIBLIOGRAPHY	110
APPENDIX 1. SPEECH MATERIALS.....	120
APPENDIX 2. TABLE OF ICT TOOLS AND PLATFORMS RELATED TO INTERPRETING TECHNOLOGY	124

APPENDIX 3. ETHICS COMMITTEE APPROVAL	126
APPENDIX 4. THESIS/DISSERTATION ORIGINALITY REPORT	127



LIST OF ABBREVIATIONS

AI : Artificial Intelligence

AIIC : International Association of Conference Interpreters

API : Application Programming Interface

AR : Augmented Reality

ARI : The Automated Readability Index

ASR : Automatic Speech Recognition

CAI : Computer-Assisted Interpreting

CI : Consecutive Interpreting

EM : Effort Model

ER : External Resources

ESIT : École Supérieure d'Interprètes et de Traducteurs (School for Interpreters and Translators in Paris, France)

ETI : École de Traduction et d'Interprétation (School of Translation and Interpreting in Geneva, Switzerland)

EVS : Ear-Voice Span

HMI : Human-Machine Interaction

ICT : Information and Communications Technology

LLM : Language Learning Machine

MFD : Mean Fixation Duration

MI : Machine Interpreting

MT : Machine Translation

NER : Named Entity Recognition

NLP : Natural Language Processing

PE : Post-Editing

RI : Remote Interpreting

RSI : Remote Simultaneous Interpreting

S2ST : Speech-to-Speech Translation

SCI : Sight-Consecutive Interpreting

SI : Simultaneous Interpreting

SMOG : Simple Measure of Gobbledygook

ST : Speech Translation

TD : Translation Dictation

TIS : Translation and Interpreting Studies

UI : User Interface

VR : Virtual Reality

WER : Word Error Rate



LIST OF TABLES

Table 1: ICT Tools and Platforms Related to Interpreting Technology

Table 2: Advantages and Disadvantages of Tablet Interpreting (Goldsmith, 2018, p.357)

Table 3: Readability Index Results and Lexical Density Ratios of Speech Materials

Table 4: Detailed Descriptions of Speech Materials (Duration, Length, Units of Meaning)

Table 5: Word-Error-Rate Results and Precision of ASR in Named Entity Recognition

Table 6: Distribution of Speech Materials per Participant

Table 7: Instances of Disfluency Markers per Participant



LIST OF FIGURES

- Figure 1.** The conceptual spectrum of interpreting drafted by Pöchhacker
- Figure 2.** Glossary creation and editing in InterpretBank
- Figure 3.** The memory feature of InterpretBank
- Figure 4.** The main interface of InterpretBank ASR
- Figure 5.** Glossary management page of Interpreter Assist
- Figure 6.** ASR Feature in KUDO Interpreter Assist
- Figure 7.** The user interface of SmarTerp
- Figure 8.** The workflow of the ASR-CAI integration in the case of InterpretBank
- Figure 9.** The functionalities of Livescribe™ Echo® Smartpen
- Figure 10.** The main layout of Sight-Terp (Tablet View)
- Figure 11.** A segmented text on the interface of Sight-Terp
- Figure 12.** Named entities highlighted in Sight-Terp interface
- Figure 13.** Digital Notepad feature of Sight-Terp
- Figure 14.** The comparable results of the preliminary test: complete renditions of meaning units in %
- Figure 15.** The comparable results of the main test: complete renditions of units of meaning in %.
- Figure 16.** The durations of the performances (in minutes and seconds)
- Figure 17.** The answers to the question “How would you evaluate your experience with the Sight-Terp tool?”
- Figure 18.** The answers to the Likert item “I think the Sight-Terp tool is easy to use.”
- Figure 19.** The answers to the Likert item “Using automatic speech recognition during the consecutive interpreting task negatively affected my performance.”
- Figure 20.** The answers to the Likert item “I think the features in Sight-Terp contributed to my consecutive interpreting performance.”
- Figure 21.** The answers to the question “Do you think the automatic speech recognition function in Sight-Terp is accurate and reliable?”
- Figure 22.** The answers to the question “Which automatically generated output did you use for support during consecutive interpreting?”
- Figure 23.** Answers to the question “Would you use the Sight-Terp tool in your future professional life?”

LIST OF CHARTS

Chart 1. The procedure followed in the study



INTRODUCTION

The key role of information and communication technologies (ICT) in interpreting is inarguably prominent considering recent tailor-made technological solutions for interpreters. Remote interpreting (RI) solutions have changed the way interpreters work and created a digital identity along with its problems and contributions. Machine interpreting (MI), on the other hand, though far from human parity, has the potential to create thought-provoking debates on user perception, multilingualism, and communicative perspective. The advancement of technology has brought about a plethora of tools and solutions to enhance the accuracy and efficiency of interpreters. With the use of computer-assisted interpreting tools (CAI) and natural language processing (NLP) applications, interpreters now have access to a whole new world of linguistic and technical possibilities, which can revolutionize the way they approach their work.

Computer-assisted interpreting is defined as software which is ‘specifically designed and developed to assist interpreters in at least one of the different sub-processes of interpreting’ (Fantinuoli, 2018b, p. 12). CAI tools emerged to fulfil the common objective, which is helping interpreters in a wide range of productivity and quality-related tasks from easing cognitive load to conference preparation and terminology organization. As a matter of fact, technological trends in the field of interpreting have changed with new developments in natural language processing, speech technologies, general artificial intelligence and changing role of interpreters with the rise of remote simultaneous interpreting (RSI) and the so-called technologization process or ‘technological turn’ (Fantinuoli, 2018b) has changed the way “computer-assisted interpreting” is perceived.

Automatic speech recognition technology is game-changer for the new generation CAI tools. The quality of ASR systems has been incrementally improved thanks to new advancements in deep learning¹, which brought about the question of whether CAI tools and ASR can be integrated. ASR-integrated CAI tools have been proposed and designed

¹ Deep learning is a subset of artificial intelligence (AI) that focuses on teaching machines to learn and process data in ways that resemble human learning.

to alleviate the cognitive strain on interpreters during the interpreting process, while simultaneously augmenting their processing capabilities. The aim in principle is to automate the querying system in real-time in simultaneous interpreting and make it possible to automatically display the reliable transcript of the source speech in a short time that fits into interpreters' ear-voice span (EVS). These new generation ASR-enhanced CAI tools have newly gained traction thanks to the tools (or projects) such as InterpretBank (Fantinuoli, 2016), SmarTerp (Rodriguez et al., 2021), VIP (Corpas-Pastor, 2021) and KUDO Interpreter Assist (Fantinuoli et al., 2022).

ASR with “considerable potential for changing the way interpreting is practiced” (Pöchhacker, 2016, p. 188) has a pivotal role in shaping the concept of human-machine interaction in the context of interpreting. Several empirical studies are questioning possible ASR implementation as an automated querying system (Hansen-Schirra, 2012; Fantinuoli, 2017), investigating the feasibility of ASR-enhanced CAI tools in the context of problem triggers (Ricci, 2020; Van Cauwenberghe, 2020; Defrancq & Fantinuoli, 2021; Rodríguez et al., 2021; Pisani & Fantinuoli, 2021; Montecchio, 2021; Prandi, 2023), using ASR for meeting the preparatory needs of interpreters (Gaber et al., 2020) and implementing ASR for supporting interpreters with the transcription of the source speech (Cheung & Tianyun, 2018; Wang & Wang, 2019). In order to enhance the depth of empirical research on CAI tools, this study deviates from the earlier studies that primarily investigated the use of ASR in simultaneous interpreting, instead focusing on the usage of an ASR and MT-enhanced CAI tool in consecutive mode. The study² attempts to fill a gap in the available literature on computer-assisted interpreting tools by proposing a prototype of an ASR-enhanced digital application and providing insights into the effectiveness of ASR and technology usage in enhancing interpreter performance in consecutive interpreting (CI), which could help shape the creation of more sophisticated CAI tools that cater to the specific needs of interpreters. The study aims at exploring and identifying a significant difference in the performances of a group of participants in CI tasks, using an ASR-enhanced CAI tool “Sight-Terp” (see section 2.4.) which is

² The scope and the results of the preliminary test of this thesis were presented with the title “Investigating the usage of ASR and speech translation in consecutive interpreter workstation: A pilot study on ASR-enhanced CAI tool prototype ‘Sight-Terp’” in the TC44 Translating and the Computer Conference organized in Luxembourg on the 22-25th of November 2022.

developed by the author within the scope of this thesis. Sight-Terp³ is a prototype of a CAI tool that initiates continuous speech recognition and provides real-time speech translation, named entity recognition and automatic segmentation of a speech. The named entity recognition (NER) function allows users to easily detect the named entities in the automated texts, such as numerals and proper names to improve their lookup mechanism. Participants' performances were tested and analyzed for accuracy and fluency using a repeated measures design. Accuracy was measured by calculating the percentage of the accurately rendered “units of meaning” (Seleskovitch, 1989) in each performance. A non-parametric statistical test (Wilcoxon Signed-Rank) was used to compare performances without technological aid and with Sight-Terp. A follow-up qualitative survey were given to the participants to obtain comparative responses and perceptions on the tool usage. The study has the potential to inform interpreting education and practice by highlighting the benefits and challenges of incorporating ASR technology in the interpreting process. By doing so, the study can also offer important practical suggestions for the future development of CAI tools for interpreters.

The first chapter of the study serves as the introduction and outlines the aim, significance, research questions, limitations, assumptions, and research definitions. The second chapter focuses on the background and the literature review of this study, starting with historical and etymological aspects of interpreting (2.1.) and cognitive dimensions of interpreting with a focus on Effort Models by Daniel Gile (2.2.). Chapter two also touches upon technology in interpreting by providing a classification of ICT tools and platforms (2.3.1). Further, in section 2.3.2., the definition of CAI tools is made with three examples of ASR-enhanced CAI tools available on the market. Section 2.3.3. then explains speech technologies in general coupled with qualitative and quantitative data from various studies on ASR integration into interpreting and translation. Section 2.3.4 mentions the usage of technological solutions for consecutive interpreting. Finally, the last section of chapter two gives a detailed description of the proposed CAI tool Sight-Terp (2.4.).

³ Sight-Terp is publicly available at: <https://www.sightterp.net>.

Chapter three outlines the methodology of the study, including its design, data collection instruments, participants, and procedure.

Chapter four presents the findings and discussions related to the accuracy and fluency differences in interpreting performance as well as comprehensive feedback of the users.

Finally, chapter five dwells on the conclusion reached at the end of the study and provides recommendations for future research.



CHAPTER 1

SCOPE OF THE STUDY

1.1. AIM OF THIS STUDY

The purpose of this study is to evaluate the effectiveness of the ASR-enhanced CAI tool Sight-Terp (<https://www.sightterp.net>), developed in the scope of this thesis, in enhancing the performance of consecutive interpreters by facilitating real-time speech translation, named entity recognition and automatic segmentation. The research aims to investigate whether the use of Sight-Terp improves the accuracy and fluency of CI as its primary objective. By means of the within-participants repeated measures design, the study seeks to empirically test the performance of a group of interpreters who will use Sight-Terp during the post-test phase. Furthermore, the research attempts to collect qualitative feedback from the participants through a follow-up survey, which will offer insights into their experiences and perspectives of using the tool. The contribution of this study to the field of interpreting will be to provide evidence of the effectiveness of ASR-based CAI tools in improving interpreters' performance by identifying a significant difference in participants' performance.

1.2. SIGNIFICANCE OF THIS STUDY

Process-oriented translation and interpreting research in experimental settings have gained traction in recent decades. Within the scope of interpreting research, research trends investigating the impact of technology-enabled interpreting tools on interpreters' tasks have mostly centred upon simultaneous interpreting. Recognizing the need to expand empirical research on computer-assisted interpreting (CAI) tools, this study diverges from previous investigations that primarily focused on automatic speech recognition (ASR) in simultaneous interpreting. Instead, it examines the utilization of an ASR and machine translation (MT) augmented CAI tool in consecutive interpreting. By proposing a prototype digital application, the study aims to address a gap in the current body of literature pertaining to CAI tools. The empirical research conducted in this study can provide valuable insights into the effectiveness of these tools in improving interpreter

performance and can inform the development of more advanced CAI tools. This thesis distinguishes itself by employing the English-Turkish language pair, whereas similar studies investigating ASR in interpreting have predominantly focused on high-resource European languages or Chinese. In addition to addressing the research questions posed by the methodology, this study seeks to compile a comprehensive table in the literature review section that highlights the various tools and platforms associated with information and communication technologies in interpreting. By doing so, it aims to provide an extensive overview of the resources that influence, either partially or entirely, the practice of interpreting.

Moreover, the methodology utilized in this empirical study might bring new questions as to whether or not we need new methodological designs for product-oriented CAI tool research for better generalizability, particularly in technology-assisted CI. The results of this study can have practical implications for professional interpreters, interpreter training programmes and speech technology developers, as they can inform the development and integration or introduction of more efficient and effective interpreting technology tools, particularly enhanced with AI and automatic speech recognition. Finally, the results of this study can lead to a better understanding of the potential of human-machine interaction in interpreting and contribute to ongoing efforts to improve the quality of interpreting services.

1.3. RESEARCH QUESTIONS

1. Does the use of the CAI tool Sight-Terp in consecutive interpreting, which provides both a source transcription and a machine translation output, lead to a significant improvement in the interpreting accuracy of interpreters compared to their performance without technological aid?
2. Are there significant differences in the number of disfluencies (pauses, hesitations, repetitions, stuttering, false starts) between pre-test performances without CAI support and post-test performances with Sight-Terp support?
3. How do users interact with the tool Sight-Terp? Do its interface design and ergonomic features meet the required standards for efficient and effective interpretation?

1.4. LIMITATIONS

The results obtained from our empirical evaluation must be interpreted in a nuanced manner, as they are subject to certain limitations. One such limitation is that the experiment of the study is conducted with students/novice interpreters. On the other hand, the language pair used in this study is Turkish and English and the interpreting task requested from the participants is in the direction from English into Turkish. The directionality is another phenomenon which may bring other factors and interferes with the accuracy and completeness of the interpreting performance particularly when it comes to technology-mediated interpreting scenarios. The use of pre-recorded speeches may not be reflective of the challenges and demands of live interpreting, which could impact the generalizability of the results. It is also critical to acknowledge that variables such as specific domains, speech characteristics, and accents - among other factors - are highly relevant and may significantly affect the tool's performance and usability. As a fourth limitation, the ASR system that Sight-Terp relies on is Microsoft Azure Speech Recognition API. At the time of writing this thesis, the Microsoft Speech Recognition API is considered to be one of the best ASR systems when compared to other equivalent software. However, this limitation should still be taken into account when evaluating the proposed software's overall performance and effectiveness.

1.5. ASSUMPTIONS

1. All participants are presumed to possess similar but comparable skills and levels of expertise in consecutive interpretation.
2. The participants are assumed to perform to the best of their ability and to be motivated to achieve high levels of accuracy and fluency in their interpretation, regardless of the presence of technological aids.
3. The participants are assumed to be honest and sincere in their self-assessment of their performance and to provide accurate responses in the questionnaires.
4. The reliability index results for the materials used in the research are presumed to be adequate and valid for assessing the validity of the performances and the pre-test and post-test speeches are assumed to have similar levels of difficulty and content familiarity.

5. The laboratory conditions are assumed to affect all subjects in a similar manner and that all subjects participate in the tasks with their utmost focus and concentration.

1.6. DEFINITIONS

Automatic Speech Recognition (ASR): ASR is a subfield of natural language processing and artificial intelligence (AI) that focuses on the development of algorithms and models to convert spoken language into written text.

Speech Translation (ST): Speech Translation is a machine learning algorithm that utilizes a variety of techniques and models to facilitate the process of translating spoken language from one language to another.

Computer-Assisted Interpreting (CAI) tools: CAI tools refer to a wide range of computer programs that have been developed with the primary purpose of supporting interpreters in one or more of the diverse sub-processes of interpreting. CAI tools provide human interpreters with real-time support in the form of speech recognition, translation, and other tools to enhance their interpretation performance, providing aid in the interpreting process.

Named Entity Recognition (NER): NER is a natural language processing task (or technique) used to identify and extract important entities such as names, locations, and organizations from a text, providing a more comprehensive understanding of the information being processed.

CHAPTER 2

THEORETICAL BACKGROUND

This chapter delineates the theoretical background of this thesis and provides a broad literature review of the core concepts that are linked with the professional, academic technological aspects of interpreting. In the first section, after a historical and etymological overview of interpreting per se, modes and settings of interpreting are defined with their precise subsections. The second section outlines the main principles of the cognitive dimension of interpreting with a particular focus on Daniel Gile's Effort Models, which are closely associated with the cognitive aspects of interpreting. Section three of this chapter elaborates on the information and communication technologies in interpreting and classifies technology-relevant interpreting tools and platforms in a single frame. Further, the brief description of speech technologies including ASR integration into the interpreting and translation are described and outlined along with relevant data from qualitative and quantitative studies. Moreover, a subsection is allocated for detailing the use of technology in CI, with a few articles published so far for a better understanding of recent approaches. Finally, the last section introduces the computer-assisted interpreting tool Sight-Terp, which this thesis is grounded on, and provides an elaborative description of its features.

2.1. INTERPRETING: AN OVERVIEW

Throughout history, interpreting was always required in any cross-linguistic communicative event in which across barriers of culture and language. It has been used for centuries to facilitate communication between individuals or groups who speak different languages, playing a crucial role in facilitating communication between people of different languages. The use of interpreters has continued to evolve and expand over periods of history. With the rise of globalization, communication between countries increased, so has the demand for interpreters. This has caused the interpreting industry to become more professional and standardized, creating professional groups and introducing interpreter training and certification programs. As a result, the ancient human practice of interpreting has undergone many social, cultural and most importantly professional

phases up until now. This section begins with a brief introduction to the concept of interpreting and its definition along with its history. It then goes on to explain the ramifications of the practice with different modes and settings.

2.1.1. Defining Interpreting

Briefly defined, interpreting is the act of transferring a message from a language (signed or oral) into another language form. Different conceptual approaches are observable in defining interpreting in a broad manner. In *Routledge Encyclopaedia of Translation Studies*, Interpreting scholar Daniel Gile defines interpreting as “the oral or signed translation of oral or signed discourse, as opposed to oral translation of written texts” (2009, p. 51).

Many languages have a corresponding equivalent word for interpreter and interpreting which are distinct from the words used for (written) translation. Etymologically, the first trace comes from the Akkadian word *targumannu* and its corresponding form *turgemana* from Aramaic the semantic component of which is ‘to explain’ (Pöchhacker, 2015, p. 198). The word finds its correspondence as *tarjuman/targuman* in Arabic, *dragoumanos* in middle Greek, *dragumannus* in middle Latin, *dragomanno* in Italian, *drugemen/drogman* in French, *tercüman* in Turkish, *tolmács* in Hungarian. The semantic inference of ‘explaining’ in these words have also a root in the greek word *hermeneus* or *hermeneuties*, referring to the Greek god Hermes interpreting the ethereal communique of the gods to the language of mortals for the sake of humanity.

The English term "interpreting" has its origins in the Latin words *interpretes* and *interpretari*. These words travelled through Old French and Anglo-French before finally being incorporated into modern English, accommodating diverse dialects and linguistic norms. As a result, the term has taken on different meanings in different contexts, with some restricting it to the act of facilitating communication between multilingual speakers and others embracing a more expansive interpretation that includes any kind of translation, whether in the form of written or spoken.

Apart from the etymological origin, it is also possible to draw a line in the distinction between translation and interpreting in that interpreting is performed ‘here and now’ and its feature of ‘immediacy’ makes the word ‘interpreting’ distinguished from other translational activities (Pöchhacker, 2016, p. 10). This denomination allows for the incorporation of other manifestations like signed language interpreting and excludes dichotomies of oral vs written translation by getting away from the common definition of “the oral translation of an oral discourse” (Gile, 1998, p. 40; 2004). Otto Kade defines the practice of interpreting as “the source-language text is presented only once and thus cannot be reviewed or replayed, and the target-language text is produced under time pressure, with little chance for correction and revision” (1968, as cited by Pöchhacker, 2016). This definition clearly articulates the feature of immediacy as the interpreter has limited potential to access the source text (can be substituted with “acts of discourse” and/or “utterances”) in its “one-time presentation” (p. 10). All in all, all definitions feature interpreting as an in-the-moment activity that focuses on facilitating oral communication.

2.1.2. History of Interpreting

Throughout history, mediation, reciprocity, connectivity, and interconnectedness have always been at the heart of the engagement of civilizations, countries, tribes etc. This engagement at the basis of all cultural interactions was wealth, reputation, invasion, and the struggle for sovereignty. Whether in conflict or not, peace-making has also been also a matter of talking and therefore of language. Having been older than the invention of writing, interpreting has taken an inevitable and crucial role in war, peace, trade, and administration in addition to its undeniable role in peace negotiations, social interactions of civilizations, the spread of religions and in the context of many periods.

Historically, records about interpreting are not in abundance for some presumable reasons, particularly prior to middle age. First, interpreting might have been considered a daily, common activity. Secondly, people in power in history writing did not consider the interpreter’s name worth mentioning, which resulted in a lack of historical documentation (Roland, 1982, p. 4). Another possible reason is the merit of invisibility as an integral ethical principle upheld by interpreters. As such, they were not considered to be worth recording in the official minutes and administrative documents. The earliest

known evidence of interpreting is from historical documents inscribing or mentioning the interpreter engaging in the practice of interpreting such as the hieroglyph from ancient Egypt depicting a communicative action between parties (Delisle & Woodsworth, 2012; p. 248) or a handful of documentary evidence on the role of interpreters in the Roman Empire (Giambruno, 2008, p. 28).

Interpreters escorted conquerors as they marched into foreign lands, assumed important roles in diplomacy and government in Ancient Egypt and in Ottoman Empire, had social privileges in many societies (Diriker, 2005, p. 88), constituted a recognized occupational group in Rome (Hermann, 2002). In ancient times, they mostly consisted of people with multiple ethnic backgrounds, slaves, or prisoners (Roditi, 1982). Correspondingly, the motivation for embarking on an expedition was not limited to religion but trade, power and annexing new areas. Conquerors selected their interpreters from the land conquered by taking them to the native country to teach their language (Andres, 2012, p.3). Ottoman interpreters, the dragomans, who were mostly in charge of embassies and consulates of European states in cities under Ottoman rule, were from non-muslims of Christian communities of Fener and Pera districts who were knowledgeable with western culture and languages (Hitzel, 1995; Abbasbeyli, 2015). This was seen in ancient Greeks, who were not eager to learn new languages as they think their language is superior and made interpreters from bilingual foreign people whom they call “barbarians” (Wiotte-Franz, 2001).

Profession-wise, it is also possible to trace the old code of ethics stipulated for interpreters. Mexican interpreters called “Nahuatlatos” were actively used in the Spanish influx into Central and South America. In this specific historical context, the striking point to lay out is that partly comprehensive legislation on interpreters was drafted by Spanish authorities which enshrines the training, accreditation, and definition of interpreters in a code of ethics (Baigorri-Jalón, 2015, p. 16). Overall, the origins of interpreting hark back to ancient civilizations. However, it wasn't until the 20th century that interpreting became a globally recognised profession, influenced by the convergence of significant political, technological, economic and social advances that played a crucial role in its development and growth.

2.1.3. Interpreting in Modern Times

The oldest and, at the same time, one of the most modern professions, interpreting has undergone many transitions on its way to institutionalization and becoming a full profession as well as an academic discipline. In the past 100 years, interpreting experienced new transformations and ramifications with new modes emerging in new settings mostly driven by economic, political, and social developments.

The widespread adoption of multilingualism in international conferences became possible after the emergence of official French-English bilingualism at the League of Nations in the early 20th century. This was a remarkable turning point in that it ensured multilingualism at international conferences and solidified the role of interpreters in facilitating communication between diverse linguistic backgrounds. Before the end of the First World War and the Paris Peace Conference of 1919, the prevalence of French in diplomatic proceedings was such that the demand for interpreters was minimal, as most participants were fluent in the language. On the rare occasion that a delegate was unable to speak French, they were assisted by a personal secretary or interpreter. Nevertheless, considering the need for interpreting was much less than in today's era of globalisation, conference interpreting was not considered a profession in its own right at the time. During these times, CI was mostly used for the meetings though it would double the duration of them. SI with equipment was not considered thoroughly until the twenties. Chronologically, in 1925, Edward Filene, a businessman, philanthropist and entrepreneur came up with the idea of simultaneous interpreting. He then appealed to Gordon Finlay, a staff member of the ILO, to conceive of a technique that could provide delegates with a method to listen to speeches via telephone. This system, called ‘the Filene-Finlay simultaneous translator’⁴ was operational using the available telephone equipment. It is known that, on June 4, 1927, the first meeting with simultaneous interpretation took place at the International Labour Conference in Geneva (Gaiba 1998, p. 3; Taylor-Bouladon, 2011). However, it can be said that there is uncertainty on the exact date and meeting where the first SI with equipment was used. While western scholars indicate that ILO was the first place, soviet historiography mentions that SI is used for the first time in the VI

⁴ The system was later named “International Translator System” by IBM in 1945.

Congress of the Comintern held in 1928 (Flerov, 2013). Another SI system was used at the International Conference on Energy in Berlin in 1930, invented by Siemens & Halske (Gaiba, 1998). Between 1920 and 1940, SI was used in some international conferences across Europe (Taylor-Bouladon, 2011) but CI was used still quite often, especially in parliamentary meetings of ILO and the League of Nations.

The start of the rich and storied history of conference interpreting dates to the successful deployment of SI in the infamous Nuremberg Trials of 1945-1946, which is considered to have marked a crucial milestone in the development of conference interpreting as a formal and respected profession. During these trials, interpreters were tasked with interpreting the speeches of Nazi war criminals, defendants, prosecutors and judges in English, French, Russian and German. Back then, Colonel Léon Dostert, the interpreter of General Eisenhower was entrusted to organize the language mediation process of the trials.

John Tusa and Ann Tusa in their book “The Nuremberg Trial” (1983) describe the event as follows:

“Colonel Dostert, the head of the translation section, had grouped his simultaneous translators into three teams of twelve: one team had to sit in court and work a shift of one and a half hours; another to sit in a separate room, relatively relaxed, but still wearing headphones and following the proceedings closely so as to ensure continuity and standard vocabulary when they took over; the third having a well-earned half-day off. The work was exacting. It needed great linguistic skills and total concentration. For many of those involved the subject matter imposed a further emotional strain. Working conditions were uncomfortable: the translators were cramped in their booths, which were even hotter than the courtroom. They spoke through a lip microphone to try to dampen their sound (the booth was not enclosed at the top) but not even the use of the microphone nor the huge headphones they wore could deaden the noise made by their colleagues. As they worked they had to fight the distractions of other versions and other languages”

The time-saving feasibility of the SI and its organized application over a long run throughout the trials was another sign of future usability of SI. It wasn't until the 1950s that simultaneous interpretation was widely implemented at the United Nations in New York. At this time, the interpreters who worked in the English booth for the Security Council gained nationwide acclaim as their interpretations were broadcast over the radio

(Taylor-Bouladon, 2011, p. 29). In later years, the system became operational using wired systems and wireless/infrared.

The International Association of Conference Interpreters (AIIC) was founded on in 1953. This occasion marked a turning point in the history of interpreting as we know it. This is because AIIC adopted a code of ethics and professional standards to regulate the working conditions of interpreters and to raise the profession's profile on the global stage, which was a great success. Alongside its birth, the AIIC also established complex administrative structures that continue to exist to this very day, with a highly centralized professional organization currently operating in Geneva.

Today, modern technology has revolutionised the field of interpreting, with interpreters relying on cutting-edge tools such as soundproof booths, wireless headsets and computer-assisted translation software to enhance their work. The industry has also become more nuanced, with specialist interpreters serving specific sectors such as finance, law and healthcare. Moreover, because of the extensive use of simultaneous interpreting which started with the Nuremberg trials, a great need for trained interpreters has arisen, leading to the creation of numerous degree courses around the world. The formal education started with the foundation of the École de traduction et d'interprétation (ETI) in Geneva and respectively with the HEC School of Interpreting in Paris which was later replaced by the Sorbonne School of Interpreting and Translating (ESIT). In time many courses and programmes have been established offering bachelor, master's and PhD degrees to prospective interpreters and help in the professionalization of the field.

The world has undergone significant changes since the early days of interpreting, and new modes and settings have emerged due to advances in technology. These changes have transformed the field of interpreting, and the next section will delve into the intricate details of these various modes and settings.

2.1.4. Modes and Settings of Interpreting

It is possible to draw a conceptual map of interpreting with different settings and constellations such as inter-social and intra-social settings and the situational

constellation of interaction (i.e., conference interpreting vs dialogue interpreting) (Pöchhacker, 2016). Among many classifications, a distinct division can be drawn based on the methods and contexts as the practice of interpreting takes place in a number of different modes and settings, each of which presents unique challenges and opportunities for the interpreter.

In the literature, there are no precise lines when it comes to classifying interpreting based on the settings, in which the action takes place, and the modes, which denote the temporal relationship of interpretation and the source message. Researchers tend to use different criteria while explaining the settings and modes of interpreting. According to Diriker (2018) interpreting can generally be classified based on the languages used in the communicational context (spoken language interpreting and sign language interpreting), the form of the interpretation (simultaneous interpreting, consecutive interpreting, whispering, sight interpreting), and the context in which the translation is performed (conference interpreting and community interpreting). Doğan (2022) adopts a particular classification. She initially outlines types of interpreting based on the method of execution with a particular focus on consecutive and simultaneous modes and then further delineates another classification based on settings where spoken and sign language mediation is needed. Accordingly, CI has subtypes such as classic consecutive, liaison interpreting, dialogue interpreting, and over-the-phone interpreting (p. 50), while simultaneous interpreting falls under the umbrella of subtypes such as TV interpreting, whispering, video-conference interpreting, sight interpreting, conference interpreting and sign language interpreting. On the other hand, the settings, namely the subjects of interpreting, are community interpreting, court interpreting, police interpreting, disaster interpreting, (Disaster Relief Interpreters, ARÇ in short), sports interpreting, healthcare interpreting, and conflict interpreting.

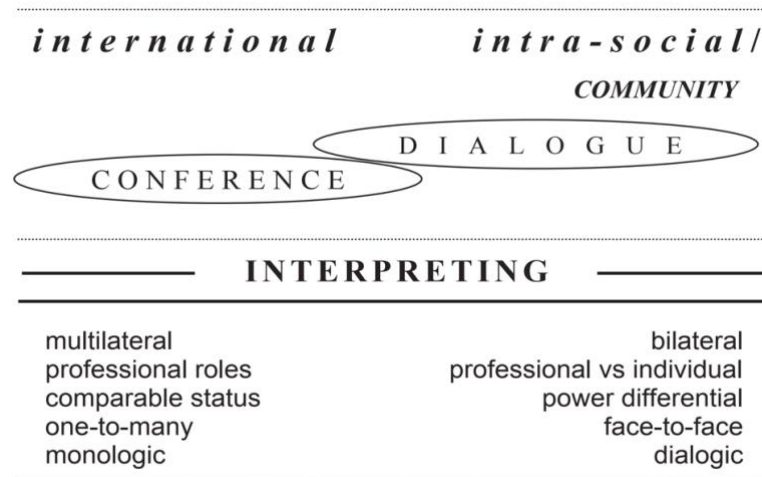
Interpreting scholar Franz Pöchhacker (2016) takes another step and creates a broader systematic typological map of interpreting based on language modality (spoken vs signed), working mode (simultaneous, consecutive, sight interpreting etc.), directionality, technology use, and professional status (professional vs non-professional). The historical prevalence of professional interpreting at international conferences and meetings has led

to the belief that conference interpreting is carried out exclusively through the use of consecutive and simultaneous interpreting. These two modes have become synonymous with conference interpreting, and it is often assumed that they are the only methods used in this type of interpreting being ‘misconstrued in a taxonomic sense’ (Pöchhacker, 2015). Pöchhacker sets forth the following interpretation regarding the topic:

Aside from the modality of the language(s) involved, which serves to contrast spoken language with sign language interpreting, the most common distinction is made in terms of the temporal relationship between the interpretation (target text) and the source text, which yields **consecutive interpreting** and **simultaneous interpreting** as the two main **modes** of interpreting. In a looser sense, different ‘modes’ can also be identified with reference to the directness of the interpreting process (relay interpreting) and the use of technology to deliver the interpretation, as in the case of remote interpreting provided in ‘distance mode’. Much more relevant, however, are conceptual distinctions with reference to the **settings** in which interpreter-mediated social contacts take place. On the broadest level, inter-social (or inter-national) scenarios, involving diplomats, politicians, scientists, business leaders or other types of representatives of comparable standing, can be viewed as different from intra-social (community-based) ones, in which one of the interacting parties is an individual speaking on his or her own behalf. The latter, subsumed under the broad heading of community interpreting, allow multiple interpreting subdivisions in terms of different institutional contexts, including legal interpreting, healthcare interpreting and educational interpreting, with numerous institution-related subtypes. (Pöchhacker, 2015, p. 199)

Moreover, by combining all distinctions, Pöchhacker details different formats of interaction by drafting the scheme in the Figure 1.

Figure 1. The conceptual spectrum of interpreting drafted by Pöchhacker (2016, p. 17)



In this subsection of the study, the main modes of interpreting, namely simultaneous interpreting, consecutive interpreting, whispering, sight interpreting and sign language interpreting will be briefly addressed. These crucial facets of interpreting will be delved into, exploring their specificities and intricacies in a succinct manner.

2.1.4.1. Consecutive Interpreting

Consecutive interpreting (CI), which is the main mode and practice used in the experiment phase of this study, is a mode of interpreting which involves listening to the speaker's message in one language with or without the use of electronic equipment, taking notes, and then delivering a full and immediate consecutive interpretation in another language. The interpreter in this mode waits until the speaker has finished a segment of speech before beginning to interpret it. In other words, the interpreter and the speaker take turns after they speak. This mode of interpreting requires the interpreter to possess a distinctive set of skills and abilities, including good memory retention, unwavering attention to detail, and note-taking dexterity, all while possessing a profound grasp of the languages in question.

CI can be practised for any duration as long as the original act of discourse continues, since the length of the speech to be interpreted is not predetermined. It involves the interpretation of both short utterances and extended speeches and thus “can be conceived

of as a continuum which ranges from the rendition of utterances as short as one word to the handling of entire speeches, or more or less lengthy portions thereof, ‘in one go’” (Pöchhacker, 2016). There are factors that affect the CI process such as the interpreter's working style, memory and situational factors. To cope with longer speeches, the note-taking technique i.e. taking notes which represent ideas and concepts rather than words is used, which was first introduced by pioneer conference interpreters in the early 1900s. Note-taking serves as a memory jogger for the interpreter. There are numerous methods and approaches in note-taking for CI each with its own unique nuances and subtleties such as mind-mapping, sentence condensation, and jotting down symbols, abbreviations, bullet points and keywords that trigger the memory of the speech content. Whether or not to use a note-taking technique divides CI into two classes: classic consecutive, where note-taking is most commonly used, and short consecutive where the duration of the speech is less than two or three minutes and does not require the interpreter to take notes.

The term 'consecutive interpreting' was considered as, so to speak, a standard or default mode of interpreting and emerged in the 1920s to distinguish it from the new method of interpreting known as 'telephonic' or simultaneous interpreting (Baigorri-Jalón, 2014; Andres, 2015), which later paved the way for the birth of the profession of conference interpreting (see section 2.1.3). Subsequent to the effective deployment of the technique of simultaneous interpreting at the Nuremberg Trial and later adoption by the United Nations, the use of CI became less widespread.

Simultaneous interpretation is commonly utilized for meetings with many languages and a big number of participants whereas consecutive interpretation is more suited for smaller sessions with technical or secret content, as well as ministerial negotiations. Additionally, CI is more flexible than simultaneous interpreting in terms of allowing the interpreter to communicate and clarify with participants, regulate the dialogical discourse, and look at the physical circumstances of the participants and their surroundings (Russel and Takeda, 2015).

Simultaneous interpreting is widely regarded as a more advanced form of interpreting and more cognitively challenging than CI. According to Gile (2001a), it is often

recommended that students begin their interpreting training with CI since it serves as a ground for the more complex task of simultaneous interpreting. Though this argument is still debatable (e.g. Seleskovitch & Lederer 1989; Russell et al. 2010), this approach can be observed in the curricula of schools of translation and interpreting around the world. Before moving on to the simultaneous mode, CI is included as one of the basic practices, along with courses in sight translation and note-taking (Niska, 2005, p. 49).

2.1.4.2. Simultaneous Interpreting

Simultaneous interpreting (SI) is a type of interpreting that involves interpreting the spoken language in real time while the speaker is still speaking. It is typically used in situations where a large group of people need to understand a speaker speaking in another language, such as at international conferences or meetings. Interpreters working in simultaneous mode are expected to produce a logically coherent output that is consistent with the source. The simultaneity of the act distinguishes simultaneous interpreting as a cognitively demanding process requiring a high level of language management.

During simultaneous interpreting, the interpreter listens to the speaker through her/his headphone and at the same time speaks into a microphone, allowing the audience to hear the interpretation via headphones or loudspeakers. This dynamic makes simultaneous interpreting demand an exceptional degree of cognitive dexterity and linguistic prowess. The interpreter must be able to keep pace with the speaker's delivery, dexterously interpreting as they listen, a feat requiring extraordinary mental agility and linguistic virtuosity.

Conference interpreters in this mode usually work in a booth where he or she can concentrate on the interpretation without distractions. Collaboration is a key aspect of simultaneous interpreting because interpreters rarely work alone. Instead, they work in pairs or even trios, each taking a 20-to-30-minute shift. This tag-team approach allows one interpreter to take a short break while the other does the heavy lifting, interpreting in real time for the audience. In this situation, teamwork is essential, with each interpreter assisting the other as needed, for example with difficult terminology. To be successful, interpreters must have an in-depth knowledge of their working languages and cultures, as

well as exceptional short-term memory. Adequate preparation is also essential, including prior research into the subject(s) of the event, which can cover a wide range of areas such as finance, medicine, law and science.

2.1.4.3. Sight Interpreting

Sight interpreting is an interpreting modality that requires the interpreter to promptly interpret written materials in real time. Sight interpreting can be considered a hybrid mode where the written source text is turned into an oral in another language. This mode has become a critical aspect of various industries like law, medicine, and professional services, where immediate verbalization of written documents or letters is imperative for the recipient's comprehension. The tempo of the interpreter's delivery in simultaneous interpreting often aligns with the pace of the speaker's speech, yet the pace of interpreting from written text lies solely within the hands of the interpreter to manipulate as they see fit. The other modes of interpreting mostly depend on auditory input, sight interpreting frees up the interpreter's memory, but also poses an added challenge in the form of allocating their processing capacity to the visual channel. The complexity of the modality is also scrutinized in the framework of translation process research. In their study, Dragsted and Gorm Hansen (2007) showed that interpreters, in sight interpreting tasks, are different than translators in temporal variables and translational approach. Eye tracking studies also show that the visual presence of the source text requires more cognitive effort and visual interference, which needs further sources allocated to cope with the lexical and syntactic complexity of a written text (Shreve, e. al., 2010).

In education, similar to the utilization of CI, sight interpreting has long been used to assess a candidate's aptitude - the ability to swiftly comprehend and articulate the core essence of a given text. This mode of testing widely considered a benchmark in determining an individual's competency in the field of interpreting (Russo, 2011). It is also commonly believed that sight interpreting can help students navigate a text in a non-linear manner and identify key information (Čeňková, 2015).

Sight interpreting has the potential to elevate the practice of simultaneous interpreting to an even greater level of accuracy and precision. This is due to the fact that conference

speeches are often written beforehand, thus granting interpreters the ability to not only listen but also peruse the text in front of them. The skills gained from practising simultaneous interpreting with written material can also be extended to the scenarios of mixed media presentations, such as presentations using PowerPoint and, most notably, presentations with real-time subtitles displayed on screens (Setton, 2015). This blend of listening and text-based interpretation results in what Gile coins as "simultaneous interpreting with text" (1995). SI with text modality is considered to be a more favourable technique, however, its intricacy is unparalleled. The written text, being dense in information and language, often lacks the fluidity and prosody of spontaneous speech. This brings the issue of the complex balance of the two acts: relying too heavily on the written text, which might result in lagging behind, and relying solely on auditory input, which can be too fast to process. The studies on this modality indicate some benefits as well. According to Lambert's study (2004), it was revealed that providing text materials to student interpreters impacts their simultaneous interpreting performance. The results indicated a substantial improvement in their performance when they were given ten minutes to prepare with the text being made available to them. Likewise, Lamberger-Felber (2001, 2003) examined the impact of simultaneous interpreting with and without text on target-text accuracy and omissions. A remarkable difference was observed in the proportion of correctly translated proper names and numbers when interpreters had access to written text, in contrast to the figures obtained in the absence of written text. The accuracy soared to 98% with time to prepare and 92% without.

It is important to note that ASR aid in consecutive interpretation, which is what this study partly aims to investigate in a product-oriented methodology, shows similarities with SI with text since the text is available as a reference during the execution of the interpreting practice. Therefore, ASR with consecutive interpreting might be entitled to *consecutive interpreting with text* or *sight-consecutive* modality.

2.1.4.4. Whispering (Chuchotage)

Whispering, also known as chuchotage, is a different form of simultaneous interpreting. Known for its use in intimate and close-quartered situations, such as business dealings or guided tours, this mode provides a one-on-one, personal experience for the listener. The

interpreter, who is physically near the listener, whispers the interpretation, ensuring unobtrusive communication without disrupting the pace of the meeting or event.

2.1.4.5. Sign Language Interpreting

Sign language interpreting involves the interpretation of verbal communication into sign language and vice versa, providing an unparalleled level of access and understanding for people with hearing impairments such as deafness and hearing loss. The role of a sign language interpreter requires a remarkable level of fluency in both sign language and spoken language, coupled with a comprehensive understanding of deaf culture and appropriate deaf etiquette in order to effectively interpret the intended message.

2.2. EFFORT MODELS IN INTERPRETING

The cognitive dimension of interpreting has garnered extensive interest from experts in various disciplines including neurology, psychology, linguistics, and cognitive science. This rich intellectual landscape has spurred numerous investigations into the fundamental cognitive processes involved in interpreting, including the pivotal aspects of listening and comprehension, production, and delivery. This section delves into how these crucial components of interpreting are explained by the interpreting scholar Daniel Gile's Effort Models (1997/2002).

The cornerstone of the interpreting process lies in the crucial stage of listening and analysis. It is here where the interpreter must attentively perceive the source text produced by the speaker and embark on the initial step of speech analysis. This involves delving into the source text to decipher its message and subsequently, finding its equivalent in the target language. Consequently, the speech is produced in a series of processes that range from the initial formation of the message in the mind to speech planning and implementation. In interpreting studies, the first modelling attempt at the translational process was put forward by Danica Seleskovitch (1962) and later developed by Lederer (1981). Seleskovitch's contribution to the cognitive analysis of interpreting is widely known, particularly for her triangular process model of interpreting. In this model, 'sense' is seen as the culmination of the process, rather than mere linguistic transcoding. More

specifically, it is the interpreter's ability to grasp and convey the underlying 'sense' of a message rather than fixed linguistic correspondences that is the essential component of interpreting. 'Sense', according to Seleskovitch, is a deliberate cognitive addition to linguistic meaning, with the added characteristic of being non-verbal. In general, the main idea in the interpretive theory is that 'deverbalised' meaning is more important in translation than linguistic conversion processes.

Later, more comprehensive multi-phase models are created focusing particularly on 'processing difficulties' (Pöchhacker, 2016). In this regard, Daniel Gile's 'Effort Models' (1985, 1997/2002) is based on the idea that in situations where cognitive decisions are necessary to complete tasks, the issue of multiple-task performance arises, as the combined cognitive demands may surpass the individual's capacity limit for processing. Gile's effort model (EM) posits that there is a finite amount of cognitive 'effort', with three basic processes competing for this resource. These processes, 'listening and analysis' (L), 'production' (P) and 'memory' (M), are essential components of the interpreting process. According to the model, all efforts require processing capacity and the sum of the three efforts must not exceed the interpreter's processing capacity, suggesting that successful interpreting requires careful management of cognitive resources. Gile introduced this model in 1985 and it remains a fundamental framework for understanding the cognitive demands of interpreting. The equation hereby can be solidified as $L + P + M < \text{Capacity}$.

In his later work, Gile expanded the model and added "Coordination Effort" (C) (management effort) and modelled simultaneous interpreting as $SI = L(\text{istening}) + P(\text{roduction}) + M(\text{emory}) + C(\text{oordination})$. The following set of formulas (Gile 1997/2002) was created in order to explain the relationship of the components. The overall processing capacity needs as the result of the sum of the individual processing capacity requirements (Pöchhacker, 2016, p. 91).

$TR (\text{Total processing capacity requirements}) = LR + MR + PR + CR$

$LA \geq LR$

$MA \geq MR$

$PA \geq PR$

$$CA \geq CR$$

$$TA \geq TR$$

Gile consequently assumes that the entire available capacity must be equivalent to or greater than the total requirements. His contribution demonstrates, based on all these formulas, that during the interpreting process, an interpreter operates within the limits of their own capacity. For the interpreting process to proceed smoothly, the available capacity for each effort must be greater than or equal to the capacity required by the relevant task. If an effort is not performed adequately, there might be errors, omissions and infelicities such as incomplete comprehension or incorrect target reformulation or incomplete retrieval of information. The EMs in SI, devised by Gile, are underpinned by the Tightrope Hypothesis (Gile, 1999). In essence, this theory posits that interpreters, much like tightrope walkers, operate on the brink of cognitive saturation (Gile, 2009, p. 198). This precarious balancing act is a constant challenge, as they must coordinate various sub-tasks. Gile's analysis reveals that when the interpreter's cognitive capacity reaches its limit, errors and "infelicities" (EOIs) occur. These missteps stem from an inability to effectively deal with "problem triggers" (Gile, 1999, p. 157), such as specialized terms, proper names, and numerical data, which demand heightened cognitive resources.

Gile has created other models that serve to represent the distinct challenges and efforts associated with various interpreting modalities, such as simultaneous interpreting with text, consecutive interpreting, sign language interpreting, and even remote interpreting. Due to its relevance to this study, I will focus on EMs in CI and EMs for human-machine intreraction (HMI) briefly below.

2.2.1. Effort Models in Consecutive Interpreting

Different from SI, in CI (with notes), the model includes other operations since different tasks are included. To be more precise, during the listening phase, the listening effort is the same as in SI but another production effort is executed when the notes are manually produced for memory-jogging. Additionally, during the listening phase again, a short-term memory effort is required to store the information until it is noted (Gile, 2001).

During the reformulation phase, three efforts are required: the note-reading effort (deciphering), the long-term memory effort, which entails retrieving information from long-term memory and reconstructing the speech content, and finally, the production effort, the operation for generating the target-language speech.

Ultimately, for CI with notes, the following model can be drafted (Gile, 2023):

Listening and Comprehension phase: **L + M + NP + C** (*NP: Note Production*)

Reformulation phase: **NR + SR + P + C** (*NR: Note Reading SR: Speech Reconstruction from Memory*)

Based on this model, it is worth noting that the interpreter is able to dedicate a greater degree of attention to the monitoring of their output during the speech, compared to the simultaneous interpreting process, where such monitoring may be more difficult to accomplish due to the demands of real-time production. Similarly, in SI, as it involves the simultaneous processing of two languages in working memory (Gile, 2001), interpreters devote some attention to inhibiting the influence of the source language to avoid ‘linguistic interference’ (p. 2), making it a more challenging task. Conversely, in CI, the effort of inhibiting the source language influence might be much weaker or even non-existent since the notes taken are shorter, more summarized and organized. Therefore, from this point of view, note-taking during comprehension would inflict more cognitive requirements whereas cognitive pressure during the reformulation phase in CI with notes would be relatively less. However, this balance may shift when technological aids like Sight-Terp are incorporated into the interpreting process. The equilibrium could alter depending on which subtask the technology helps to reduce cognitive load for.

2.2.2. Effort Models in Human-Machine Interaction

Daniel Gile suggests in his keynote speech (2020) that Effort Models could give rise to new versions if researchers and teachers discover novel functions connected to significant attentional resource requirements in interpreting. A potential situation could arise if interpreters were required to direct significant attention to interaction with more screens, interfaces, and technological tools. A recent development might serve as an example.

During the COVID pandemic, remote interpreting platforms grew in number and for the last three years, there has been an increasing volume of demand for interpreters working remotely. When team members are not in the same location, the communication between boothmates must be through video-conference platforms which, though with some essential similarities, have different interfaces and functions. In similar veins, CAI tools (see section 2.3.2), especially those designed specifically for in-booth scenarios, have certain functionalities that require familiarity and additional cognitive resources. In this respect, Gile (2020; 2023) postulates the following model (for SI), taking into account the changing technology and working environments:

SI: R + M + P + HMI + C

Here, 'R' is for reception, 'which can be both auditive and visual' (Gile, 2020) while HMI stands for human-machine interaction. HMI is a broad concept which might have different efforts. In the example of remote interpreting, Gile adds the turn-taking effort as TT. Turn-taking in remote interpreting can be more complex and challenging than in traditional settings, due to factors such as latency, audio quality, and coordination with other participants. However, in general, there are many combined efforts required to manage and troubleshoot technology-related issues, such as connectivity problems, audio and video settings, and platform-specific features.

In the main study of this thesis, the software Sight-Terp uses ASR to generate the speech transcript with which the interpreter can deliver the interpretation by looking at the script. Since there's no need for note-taking⁵, the interpreter can allocate their attention to every detail in the speech, and focus on formulating the interpretation in their mind, without the extra cognitive pressure of note-taking. Though that would mean less cognitive pressure on the comprehension phase, the software's constant visual presence of the auto-generated text may induce more cognitive pressure on the reconstruction phase, requiring the interpreter to reformulate and adjust their interpretation constantly. The additional features of Sight-Terp (named entity highlighting, automatic segmentation), which is

⁵ Sight-Terp, in fact, allows for digital note-taking with a stylus (like Apple Pen). Though the feature of digital note-taking embedded in Sight-Terp is described in the study, note-taking is excluded from the main study and the participants are instructed to only use the ASR function.

detailed in the following sections (section 2.4), are deployed in order to mitigate linguistic interference which is more generally associated with sight interpreting (Agrifoglio, 2004).

Based on Gile's Effort Models, a formula for an effort model specific to *sight-consecutive* interpreting can be drafted. In *sight-consecutive* interpreting, the interpreter relies on a text-based reference generated by an ASR system, which reduces the cognitive load associated with listening and memory to some extent. Consequently, the effort model for sight-consecutive interpreting might place more emphasis on the analysis of the text, the production of the target language, and the coordination of these efforts. In light of these restrictions, mitigations and possible cognitive requirements brought by Sight-Terp the following model can be drafted to encompass *sight-consecutive (SCI)* modality:

SCI: Listening and comprehension phase: $L + M + NV + C$

(**L**: listening, **M**: memory, **NV**: note verification, **C**: coordination)

NP is replaced with NV (note verification) implying the effort of the interpreter to monitor the accuracy of the ASR, make corrections and take up strategies and coping mechanisms accordingly. The cognitive demands of using the tool will likely vary depending on the quality of the ASR output.

Reformulation phase: $BR + SR + P + C$

(**BR**: bilingual note reading, **SR**: speech reconstruction, **P**: production, **C**: coordination)

In the reformulation phase, BR (bilingual note reading) is included to manage the bilingual format of the text MT and auto-generated source transcript, SR (speech reconstruction) to reconstruct the meaning of the source text, P (production) to produce the interpretation, and C (coordination) to manage the use of the tool. In the reformulation phase, Strong C and P are needed because of the linguistic interference potentially resulting from the bilingual format of the text MT and auto-generated source transcript together (see 2.4.1.1.).

2.3. TECHNOLOGY AND INTERPRETING

As technological advances are overhauling the interpreting sphere, they cause a shift in the traditional practices of interpreters. The proliferation of large language models (LLMs), machine translation, speech recognition technologies and other cutting-edge tools have the potential to transform the interpreting process and demand a change in the way interpreters approach their work. The impact of technology on interpreting is multifaceted. On the one hand, technological innovations have streamlined information access and work management for interpreters, leading to an increase in productivity. On the other hand, the emergence of new technologies has disrupted the demand for interpretation services in the marketplace and has overhauled the entire landscape of the industry.

The following section explores the proliferation of technology and its impact on interpreting and the latest technological developments and concepts with a particular focus on ASR-enhanced CAI tools. I will delve into speech technologies and their impact on both written translation and interpreting and examine ASR-enhanced computer-assisted interpreting tools. As technology continues to shape the landscape, I will explore the ways in which it affects consecutive interpreting, highlighting innovative methods and techniques. Finally, I will focus on the proposed tool 'Sight-Terp' and provide an insight into its intriguing features and capabilities.

2.3.1. The Emergence of Information Technologies in Interpreting

Information and communication (ICT) tools have been a driving force in the pursuit of improved quality and productivity in both translation and interpreting over the last two decades. Interpreting has not experienced such a significant impact in contrast to the transformative effects that ICT has had on translation. However, it is possible to say that there have been crucial technological advances in the field of interpreting. When discussing the evolution of interpreting in light of the emergence of information technologies, it is worth highlighting some key breakthroughs in the field. One such example can be, as stated in section 2.1.3, the advent of simultaneous interpreting. SI stands out as the first game-changing innovation which took place in the 1920s when IBM

made a ground-breaking breakthrough in developing a hardwired system for instantaneous speech transmission. Gaining popularity in several other international conferences, the wired system eventually made its mark in history by becoming an irreplaceable asset during the Nuremberg trials. Needless to say, this breakthrough changed the way interpretation is facilitated on daily basis and created an imminent social status for interpreters. The second and most important breakthrough is the introduction of the world wide web, which has revolutionized the way that interpreters access and share information, opening up new avenues for research and collaboration. The significance of the internet lies behind the crucial need for preparation for interpreting assignments: conference interpreters are constantly engaging in different “specific terms, semantic background knowledge and context knowledge” in each assignment they are in (Rütten, 2016). The World Wide Web, with its fast ability to gather information from a multitude of sources, has a powerful advantage. By streamlining the information management process, interpreters have increased the efficiency of their preparation.

Today, the current landscape of interpreting technology is a vast and varied one, characterised by a wide range of technological solutions that have played a significant role in ushering in a ‘technological turn’ (Fantinuoli, 2018b) in the profession and creating bespoke and non-bespoke computer-assisted interpreting tools. The categorization of the recent technologies of today’s interpreting technology sphere would be a line between the purpose and functions of such tools. Considering that the interpreting technology is a vast umbrella term, classification is a must for a thorough understanding indeed.

2.3.1.1. Categorization of Technologies in Interpreting

There are a couple of approaches when it comes to the classification of ICT tools in interpreting. Fantinuoli (2018a) suggests two classifications: setting-oriented technologies and process-oriented technologies. Setting-oriented technologies “primarily influence the external conditions in which interpreting is performed” (2018a, p. 155). On the other hand, process-oriented technologies include a variety of tools, such as “terminology management systems, knowledge extraction software, and corpus analysis

tools“ (p. 155), all of which aim to assist interpreters in different sub-processes and various phases of an assignment.

In parallel with this approach, according to Braun (2019), interpreting technology can be categorized into three. The first category is “technology-mediated interpreting” which encompasses all technologies employed to expand the reach and effectiveness of interpreting services, including remote simultaneous interpreting (RSI) equipment. In broad terms, technologies mediating interpreting entail distance interpreting technologies, which cover “a whole range of technologically different setups” (Ziegler & Gigliobianco 2018, p. 121). Remote interpreting can be defined as the utilization of various instruments of ICT to enable interpreter-mediated communication from a physically removed location. During the COVID-19 pandemic, remote interpreting served as the catalyst for the development of a fresh generation of conference interpreter profiles, a location-independent alternative to traditional conference settings. Moreover, the proliferation of video conference platforms (e.g., Zoom, Interactio, KUDO, and Interprefy) during the pandemic paved the way for computer-assisted interpreting tools explicitly developed for incorporation in RSI scenarios (see *Interpreter Assist* in section 2.3.2.2.). The incorporation of cutting-edge augmented reality (AR) innovations, including the deployment of advanced virtual reality goggles, can be the next evolutionary leap in remote interpreting by mitigating “the feeling of isolation” (Ziegler & Gigliobianco 2018, p. 136) and/or integrating the CAI tool interfaces on the virtual reality screen worn by the interpreter⁶ (Gieshoff, 2022).

The second category is technology-generated interpreting, which implies machine interpreting (MI) or speech-to-speech translation. MI can be characterized as a technological advancement enabling the conversion of spoken language into another language through computer programming (speech technologies)⁷. MI involves a multi-

⁶ At the time of writing, a group of three scholars at Zurich University of Applied Sciences are examining whether augmented reality technology can provide assistance to interpreters in their additional exertion of having to consult terms. In other words, the research focuses on integration of ASR-enhanced CAI tool interface on augmented reality screen by postulating that instead of switching between different types of visual information and redirecting the visual attention for CAI output, interpreters can benefit from the output directly on their augmented reality interface by wearing virtual reality headset.

⁷ The section 2.3.3. briefly focuses on aforementioned speech technologies.

step approach that generates an audible version of the translated text by creating a synthetic speech in the target language. In cascade systems, the steps are as follows: ASR transcribes oral speech into written text. This is followed by machine translation, and finally, text-to-speech synthesis is used to generate an audible version of the translated text.

The third category is “technology-supported interpreting”, which entails all technologies that can be used to augment or facilitate interpreters' preparation, performance, and workflow. In this context, technologies supporting interpreting can be considered as a wide group of technological applications and hardware that are used before, during and after the interpreting process, thereby affecting the cognitive processes behind the actual task of interpreting. CAI tools (see 2.3.3.) and other technologies that aim to enhance the performance of the task can be listed under technology-supported interpreting. The CAI tools falling under the technology-supported interpreting class has also classifications namely ‘generations’ depending on their purpose, feature and release date, as described in 2.3.3.

Drawing inspiration from Ortiz and Cavallo's list of ICT tools for interpreting (2018, p. 17), which categorizes tools by their function, specificity, and update date, I have expanded the list to include new categories such as Speech Bank, Audio and Video Conference platforms, Machine Interpreting and Real-time Speech Translation. In table 3 below, the tools have been matched according to their specificities, purposes, modalities, and features to provide a comprehensive overview of the range of tools currently available to interpreters as of January 2023. The categories are training platform, speech bank, glossary management, corpora building, terminology extraction, speech recognition, note-taking, virtual booth service, audio and video conference, machine interpreting, and real-time speech translation.

The tools under the category of ‘interpreter training’ and/or ‘speech bank’ show various platforms and software that facilitate lexical and terminological searches for both novice and expert interpreters. These tools aim to help interpreters hone their interpreting skills and strengthen their grasp of both their native language and foreign languages by allowing them to conduct deliberate practice using speech and other materials. Glossary

management, corpora building and term extraction tools (regardless of their specificity for interpreters) indicate the resources that can be used to aid interpreters during preparation, allowing them to delve deeper into the primary topic they will be interpreting. Additionally, interpreters can develop and reference personalized glossaries throughout the interpretation process, while also familiarizing themselves with the speakers' accents and backgrounds by watching videos and scouring online sources. The categories of Speech Recognition, Real-time Speech Translation, Note-taking and Virtual Booth Service include tools that are utilized for the interpreting process itself. The tools under this class are ASR-enhanced CAI tools for SI, speech translation solutions for various purposes, and note-taking applications that can be used for interpreting scenarios. Therefore, this class of categories as well as categories related to preparation/terminology can be listed under the division of technology-supported interpreting.

Platforms, where remote simultaneous interpreting can be carried out⁸ (corresponds to technology-mediated interpreting), are listed in the category of 'Audio and Video Conference'. Finally, tools under the category of 'Machine Interpreting' (speech-to-speech interpreting) are specified as 'replacement' (corresponds to technology-generated interpreting) referring to full automation of the interpreting process, resulting in a complete replacement of human interpreters. In this category, available devices and tools on the market are added based on their availability. The columns of the table show specificity (whether it is designed for interpreters), purpose (main aim of usage), modality (simultaneous interpreting and/or consecutive interpreting), and feature (remote interpreting platform, ASR-enhanced or fully ASR-powered, replacement by MI).

⁸ Platforms do not offer solutions only for conference settings but for community settings too.

Table 1. *ICT Tools and Platforms Related to Interpreting Technology*

ICT tools in Interpreting											
Categories: Training Platform, Speech Bank, Glossary Management, Corpora Building, Terminology Extraction, Speech Recognition, Note-taking, Virtual Booth Service, Audio and Video Conference, Machine Interpreting, Real-time Speech Translation											
Name	Category (main function)	SPECIFICITY	PURPOSE			MODALITY		FEATURE			
		Specific for Interpreters	Interpreter Training	Prep. (Corpora Building)	Prep. (Terminology Management)	Simultaneous	Consecutive	RI Platform	ASR	(RI) Advanced Booth Controls	Replacement
Melissi KOSMOS	Training Platform	Y	X			X	X				
InTrain	Training Platform	Y	X			X	X				
Linkinterpreting	Training Platform	Y	X			X	X				
Interpreter Training Resources.eu	Training Platform	X	X			X	X				
InterpreterQ Media Player	Training Platform	Y	X								
Speechpool	Speech Bank	Y	X								
ORCIT	Speech Bank	Y	X								
EU DG -SCIC Speech Rep.	Speech Bank	Y	X								
Interplex UE	Glossary Management	Y			X	X					
VIP Voice-text Integrated System for Interpreters	Glossary Management	Y	X		X	X	X		X		
InterpretBank	Glossary Management	Y			X	X					
KUDO Interpreter Assist	Glossary Management	Y			X	X			X		
Interpreter's Help	Glossary Management	Y			X	X					
Flashterm	Glossary Management	N			X						
Intragloss	Glossary Management	Y			X	X					

BootCaT	Corpora Bulding	Y		X							
SDL Multiterm Extract	Terminology Extraction	Y			X						
Simple Extractor	Terminology Extraction	Y			X						
Sketch Engine	Terminology Extraction	Y			X						
Terminus	Terminology Extraction	Y			X						
TermSuite	Terminology Extraction	Y			X						
InterpretBank ASR	Speech Recognition	Y				X			X		
Dragon NS	Speech Recognition	N				X	X				
Evernote	Note-taking	N					X				
Cymo Note	Note-taking	Y					X		X		
Sight-Terp	Note-taking	Y					X				
Neo SmartPen	Note-taking	N									
Livescribe Smart Pen	Note-taking	N					X				
Nebo	Note-taking	N					X				
Bamboo Paper	Note-taking	N					X				
Noteshelf	Note-taking	N					X				
Notability	Note-taking	N					X				
Penultimate	Note-taking	N					X				
LectureNotes	Note-taking	N					X				
CymoBooth	Virtual Booth Service	Y							X	X	
SmarTerp	Audio and Video Conference	Y						X	X	X	
GreenTerp	Audio and Video Conference		X					X		X	
KUDO	Audio and Video Conference							X		X	
Converso	Audio and Video Conference							X		X	
Olyusei	Audio and Video Conference							X		X	
Interactio	Audio and Video Conference							X		X	
cAPPisco	Audio and Video Conference							X		X	
VoiceBoxer	Audio and Video Conference							X		X	

Interprefy	Audio and Video Conference							X		X	
QuaQua	Audio and Video Conference							X		X	
Akkadu	Audio and Video Conference							X		X	
Catalava	Audio and Video Conference							X			
CymoMeeting	Audio and Video Conference							X			
Lingolet	Audio and Video Conference							X			
Ablioconference	Audio and Video Conference							X		X	
InterpretCloud	Audio and Video Conference							X			
TranslitRSI	Audio and Video Conference							X			
WordSynk	Audio and Video Conference							X			
Zoom	Audio and Video Conference							X			
Qonda	Audio and Video Conference							X			
Rafiky	Audio and Video Conference							X			
Ouispeak	Audio and Video Conference							X			
WebSwitcher	Audio and Video Conference							X			
ZipDx	Audio and Video Conference							X			
WebEx	Audio and Video Conference							X			
KudoAI	Machine Interpreting					X			X		X
KudoAI	Real-time Speech Translation					X			X		X
Travis Touch Go (Device)	Machine Interpreting						X		X		X
iTranslate Voice	Machine Interpreting						X		X		X
LingvaNex Pocket Translator	Machine Interpreting						X		X		X
Wordly	Real-time Speech Translation					X	X		X		X
Lingolet One (Device)	Machine Interpreting					X			X		X
Skype Translator	Real-time Speech Translation					X			X		X

One Mini (Device)	Machine Interpreting						X		X		X
CheetahTALK (Device)	Machine Interpreting						X		X		X
SSK Translator (Device)	Machine Interpreting						X		X		X
iFLYTEK Translator (Device)	Machine Interpreting						X		X		X
Waverly Labs. Ambassador (Device)	Machine Interpreting						X		X		X
Vasco (Device)	Machine Interpreting						X		X		X
Google Pixel Buds (Device)	Machine Interpreting						X		X		X
Fujitsu Healthcare Interpreter (Device)	Machine Interpreting						X		X		X
Timekettle WT2 (Device)	Machine Interpreting						X		X		X

2.3.2. Computer-Assisted Interpreting Tools

The term computer-assisted interpreting tools is defined by Fantinuoli as “all sorts of computer programs specifically designed and developed to assist interpreters in at least one of the different sub-processes of interpreting” (2018a, p. 12). CAI in some scholarly works is used to encompass all supporting technologies applied in the interpreting process. However, some authors prefer to put a distinctive line between types (bespoke or non-bespoke) of CAI tools. For example, Will's (2020, p. 47) definition of non-bespoke tools that target pre- and post-process phases is "secondary CAI tools". Such tools refer to computer-based applications employed for searching, compiling, and documenting terminologically relevant structures whereas “primary CAI tools” is designed to specifically meet the ergonomic and cognitive demands of interpreters during the interpreting process (2020). As a third category, Will postulates that software which is capable of functionalities of primary as well as secondary CAI can be referred to “integrated CAI tool” (2020, p. 48).

Fantinuoli (2016, 2018a) categorizes CAI tools into generations based on their architecture and functions. The first-generation CAI tools, designed to help interpreters manage terminology before the booth phase, are simple entry structures that offer basic look-up functionalities for glossaries. These tools offer a simpler entry structure that is more appropriate for interpreters' terminology work and offer querying in an “interpreter-friendly manner” (2018a, p. 164). On the other hand, second-generation CAI tools, in addition to basic terminology management, provide features for organizing textual material and retrieving information from corpora and other resources. Furthermore, the second generation of tools also introduces bespoke functionalities specifically designed for the in-booth phase like automating and speeding up the querying with dynamic search capabilities (2018a, p. 166).

Interpreters (particularly during SI) retrieve specialised terms from memory, which can be mentally taxing if not consolidated. They search for target language equivalents in their glossary or online databases. However, these processes are time-consuming and may distract from the task at hand. Interpreters must also contend with "problem triggers" (Gile, 2009, p. 157) such as acronyms, specialized terms and numbers, which are highly problematic and associated with increased error rates. In this context, the incorporation of glossary management is a common occurrence and asset in the existing CAI tools, and it aims to assist interpreters in various ways. It is quite versatile as it is used before, during, and after the event by offering a database of resources, tools for retrieving, memorizing, and extracting terminologies, offering multilingual glossaries and the ability to search for equivalents in the target language using a variety of database. Intragloss⁹, Interpreter's Help¹⁰, Flashterm¹¹, Terminus¹², Lookup¹³ and Interplex¹⁴ are among the tools for terminology purpose (mostly first generation CAI). Another CAI tool InterpretBank¹⁵, as a combination of first and second-generation tool, assists its users in the extraction of terminologies from preparation documents and applies a “Corpus-driven Interpreter

⁹ <http://intragloss.com>

¹⁰ <https://interpretershelp.com/>

¹¹ <https://www.flashterm.eu/>

¹² <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl?lInt=En>

¹³ <http://www.lookup-web.de/>

¹⁴ <http://www.fourwillows.com/interplex.html>

¹⁵ InterpretBank is described in detail in section 2.3.2.1.

Preparation” (CDIP) which aims at “turning the preparatory phase into a discovery-oriented task for terminology and knowledge acquisition” (Fantinuoli, 2017b, p. 29). CDIP automates the construction of the corpus that interpreters can utilize to extract specialized vocabulary in advance of an interpreting task, thereby streamlining their preparation efforts while being able to analyse the term in its context.

Thanks to the latest technological advancements, with the integration of artificial intelligence (AI) into CAI tools, third-generation CAI tools have emerged, providing a comprehensive solution for optimizing every aspect of the interpreting workflow. With glossaries created automatically from interpreters' material, the preparation task can be automated. Moreover, the amalgamation of AI and ASR technology automates the query of interpreters' glossaries, allowing for real-time support like a boothmate for all major classes of "problem triggers" (Gile, 2009). In other words, ASR-enhanced CAI tools display the target language equivalent of specialised terms, acronyms, named entities and numbers for the interpreter in real-time and enhance performance (Fantinuoli, 2017). Section 2.3.3.2. focuses on ASR integration in interpreting through CAI tools and other possible utilization of this promising technology.

It appears that interpreting technology has yet to achieve a state of perpetual professionalized expansion, wherein businesses are actively involved in creating and promoting cutting-edge and tailor-made solutions. Instead, most solutions emerge from academic research. For example, CAI tools InterpretBank and LookUp emerged from doctoral research (i.e. Fantinuoli, 2012). While these tools are based on previous interpreting research and theory, they often lack dedicated data collection to address community needs for design purposes (Frittella, 2023, p. 61). As a result, the design of these tools often mirrors the concepts and practices of the developer, who is typically an interpreter, rather than the needs of the interpreting community (Fantinuoli 2018a, p. 164).

CAI research is a relatively new field with little consensus on its specific definition. With a certain technological turn in the area of CAI and its tools, professionals and educators have started to show interest in it. The past decade has seen an upsurge in research in this

field, with some researchers like Fantinuoli (2018b) noting that it is far underutilized in an empirical manner. There is a limited number of studies conducted on the use of CAI tools in controlled environment settings. Fantinuoli argues that in order to really understand the pros and cons of CAI tools, how they affect the interpreting process, and whether interpreters can execute more effectively with their assistance or otherwise, research needs to be carried out not only using naturalistic methods (such as corpus analysis) but also under the most stringent experimental conditions (2018a, p. 170). Research in the literature varies according to its subject of analysis: evaluation of the tools' performance (e.g. Fantinuoli, 2017; Fantinuoli et. al, 2022), evaluation of the interpreters' (users) performance and user perception (mostly post-experiment questionnaires). Due to its relevance to this study, I will particularly focus on CAI tool research on users' performance.

The first group of experimental studies focus on the use of CAI in terms of effective terminology look-up and rendition accuracy with the intention of answering either exploratory or experimental research questions regarding the possible positive or negative effects of utilizing a CAI tool during the preparation or interpretation phases on the interpreters' output. Gacek (2015) and Biagini (2015) assess the efficacy of a CAI tool during simultaneous interpreting as compared to paper glossaries. While Gacek focused on gathering qualitative data, Biagini collected product-oriented data through statistical analysis of the transcriptions interpretations of the participants in the test. Both studies used the CAI tool InterpretBank, which is detailed in the following sub-section. Results demonstrated that querying the glossary with InterpretBank yielded higher terminological precision and fewer omissions than with a paper glossary. With a methodology encompassing both qualitative and quantitative data, Prandi (2015) investigated the students' approach to CAI and analysed terminological precision in students' performances. She confirms the feasibility of integrating CAI tools into interpreters' training (p. 56). Most of the studies in the CAI tools context (like this study at hand) examine the CAI tool efficiency in a product-oriented manner unlike Prandi's (2023) study at the University of Mainz. Prandi, in a cognitive line of the research, investigated in-process CAI tool use among nine interpreting students. The study implemented a mixed-method and multi-method approach that enabled a comprehensive evaluation of

the participants' cognitive capacities across three tools. In a series of experimental tasks, the students were asked to interpret three speeches from English into German using each of the three distinct tools: “a digital glossary in PDF format, a CAI tool with manual look-up, and a mock-up CAI tool with integrated ASR for terminology” (p. 166). Their gaze data and deliveries were recorded and analysed to provide nuanced insight into the effect of CAI tools on cognitive load and attention allocation during simultaneous interpreting. The research examined various performance, behavioural, and subjective metrics to create a holistic picture of human-machine interaction in technology-mediated interpreting scenarios. Moreover, with its mixed-method approach, the study brings about new methodological implications for future computer-assisted simultaneous and consecutive interpreting research. In addition to the previously mentioned research on CAI efficiency in terminology usage and lookup mechanism, a number of investigations have been conducted on ASR-supported CAI (third-generation) technologies. For the results and discussion of such research see section 2.3.3.2 which provides an exposition of the scholarly work on the feasibility of incorporating ASR technology into interpreting.

In this subsection, three CAI tools will be briefly introduced: InterpretBank, Interpreter Assist and SmarTerp. Due to its relevance to this study, only ASR and AI-enhanced CAI tools (third generation) are chosen to be outlined. The tools are available at the time of writing and these bespoke tools' ASR capabilities are fundamentally designed for supporting interpreters during simultaneous interpreting. Sight-Terp, on the other hand, uses a novel approach to integrate ASR in the consecutive interpreting workstation. Following the information about the tools, the main features of Sight-Terp will be elaborated on in the following sections.

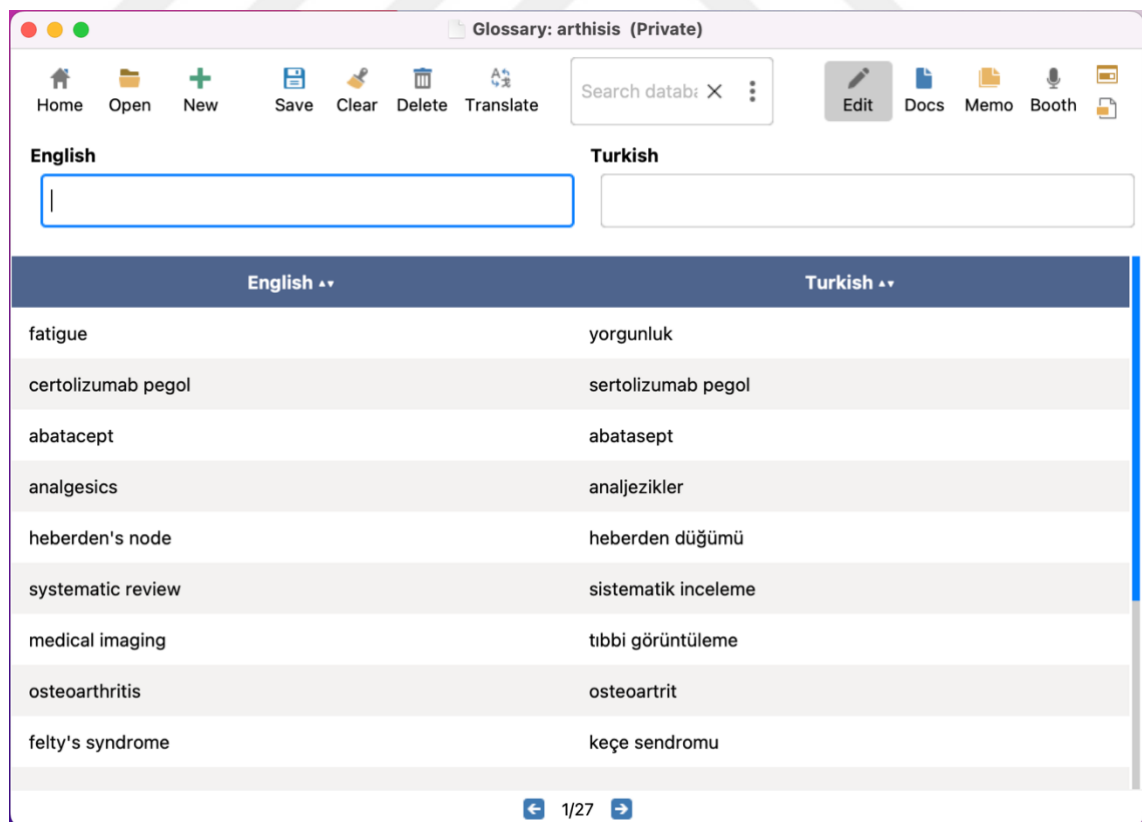
2.3.2.1. InterpretBank

InterpretBank¹⁶ is a CAI tool that is specifically designed to facilitate the work of professional interpreters by providing them with a range of advanced features and functionalities (Fantinuoli, 2009, 2012, 2016). The multifaceted InterpretBank software merges the central Edit Modality with three distinct modules that cater to varying stages

¹⁶ <https://interpretbank.com/>

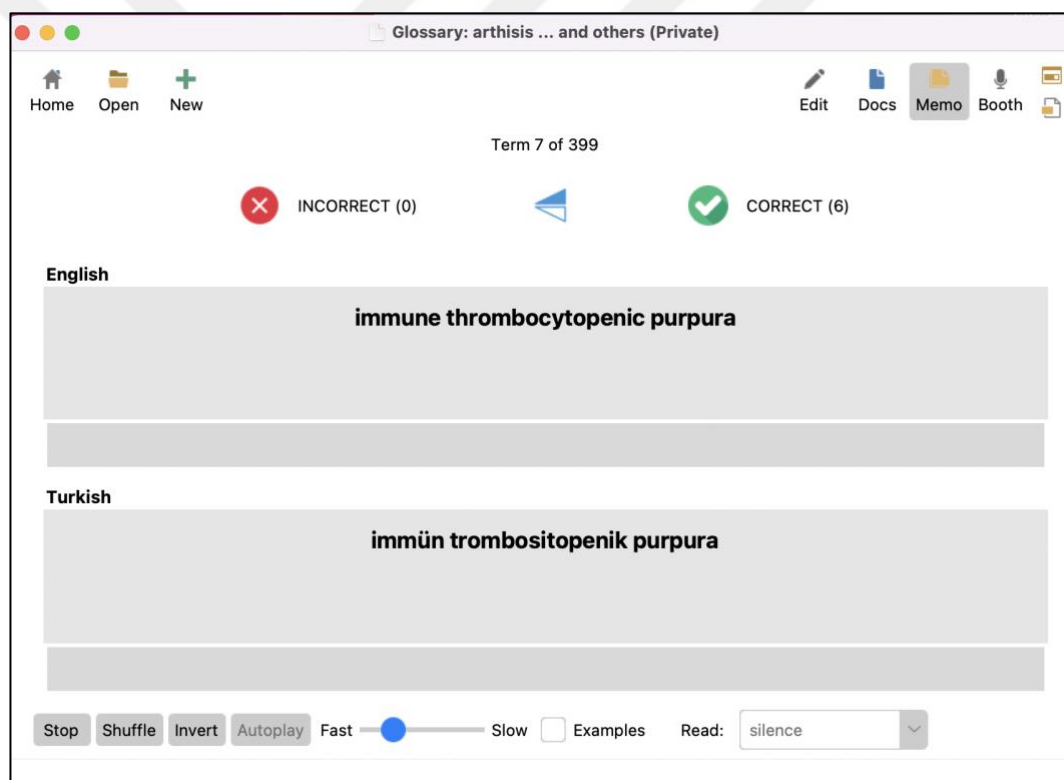
of conference preparation. With the Document Modality, Memo Modality, and Booth Modality, users can effortlessly access the database, update glossaries, merge existing ones, and transport terms between them. The integrated resources enable automatic searches for translation equivalents and definitions both online and offline. With its latest version (InterpretBank 8), users can use the AI tools feature and create automatic multilingual glossaries from a single specialized term or a selected document. This feature allows interpreters to create an instant corpus and glossary in cases when the preparation materials are not enough. Although the user must evaluate the outcomes of the automated translation, this feature streamlines preparation under tight schedules, allowing interpreters to devote their attention to higher-level processing rather than manually compiling glossaries. These functions are especially advantageous when interpreting with limited preparation materials for terminology extraction and topic identification. Although glossaries are contained within a single database, sub-glossaries and groups can be created with tags, allowing for additional structure and customisation.

Figure 2. Glossary creation and editing in InterpretBank



Interpreters, when given preparation documents or online resources, can utilize the Document Modality to optimize their usage. By creating virtual flashcards from glossaries, the Memory Modality can aid in memorizing event-specific terminology. It allows users to opt for manual or automatic presentation modes. The Booth Modality, which covers the actual interpretation phase, completes the tool's architecture. Interpreters can activate multiple glossaries in the Booth Modality, which can be queried simultaneously during the interpreting process. Furthermore, InterpretBank can search through the entire database or external resources as an emergency strategy (Prandi, 2023, p. 40).

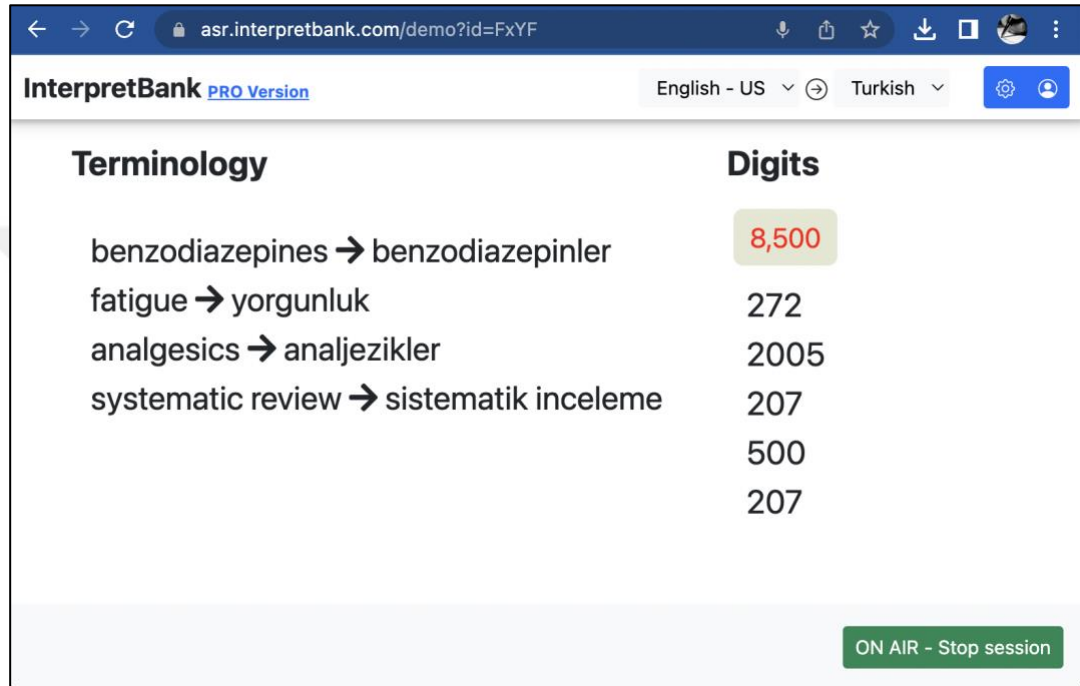
Figure 3. The memory feature of InterpretBank



InterpretBank presents an advanced feature that integrates automatic speech recognition technology to be used in the booth (InterpretBank ASR or Artificial BoothMate). This experimental feature in the freelance version represents the next step in supporting interpreters through technology, as it offers a unique approach to dealing with "problem triggers" (Gile, 1999, 157) that are typically difficult to interpret, including numbers, specialised terminology, and named entities. The ultimate goal is to create a computer-

assisted interpreting tool that functions as an artificial boothmate, running an AI-enhanced automatic speech recognition session, and providing interpreters with real-time visual support for specific terminology and numerical items during the interpreting process.

Figure 4. The main interface of InterpretBank ASR



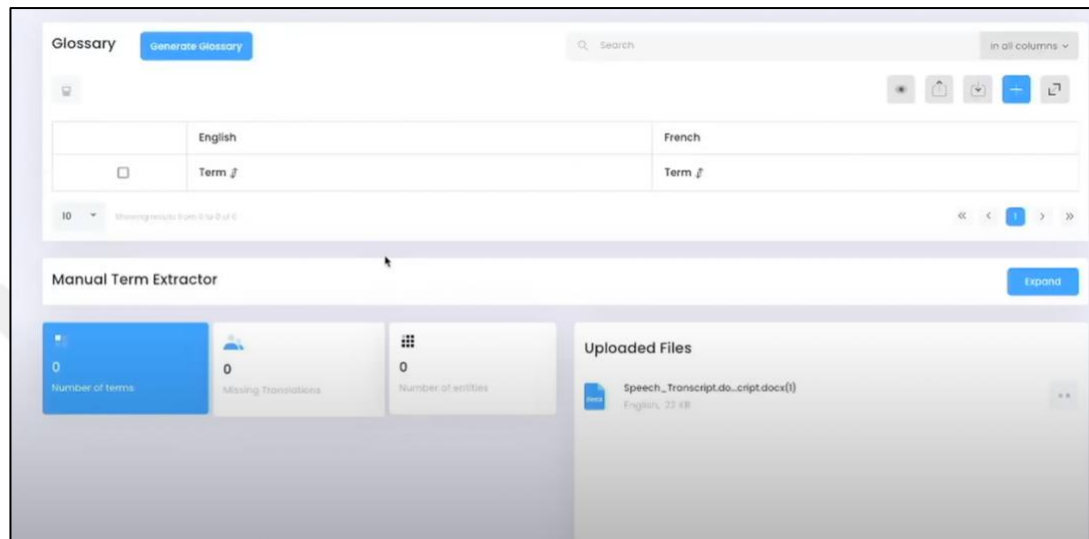
2.3.2.2. Kudo Interpreter Assist

KUDO Interpreter Assist is a product launched by Language-as-a-service (LaaS) platform KUDO¹⁷. Interpreter Assist is specifically designed for RSI scenarios, aiming to streamline the preparation process and improve rendition precision in specialized events. The tool has two main features: “an automatic glossary creation tool and a real-time suggestion system” (Fantinuoli et. al, 2022). The automatic glossary creation tool generates multilingual resources, while the real-time suggestion system offers support during interpretation sessions by providing suggestions for terms, numbers, and proper names (p. 3).

¹⁷ <https://kudoway.com/articles/introducing-kudo-interpreter-assist/>

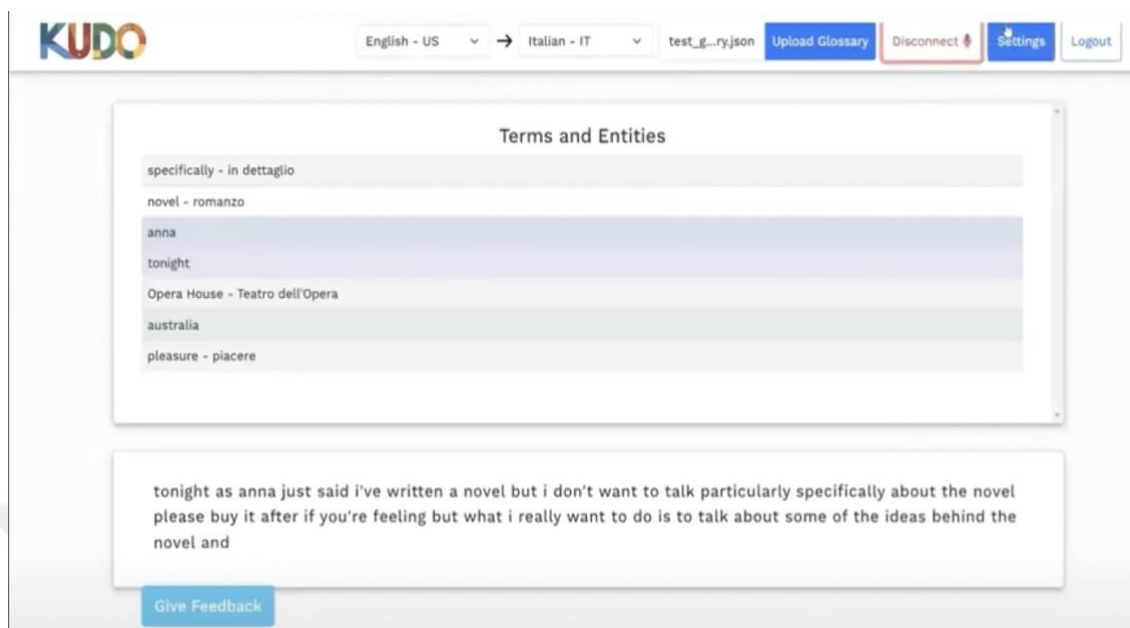
The process of creating an automatic glossary entails generating a mono or multilingual domain-specific corpus, extracting monolingual terminology, translating the terms into target languages, and refining baseline models by utilizing the produced resources.

Figure 5. Glossary management page of Interpreter Assist (kudoway.com)



The software's ASR feature is based on a cascading system of three main components (Fantinuoli et. al, 2022, p. 4). First, ASR transcribes speech in real-time and can be fine-tuned using project-specific data to increase precision. Next, a language model (LM) identifies units of interest, with terminology matched using a generated and edited multilanguage glossary, while numerals and proper names are recognized using NER. Finally, the results are sent to the interpreter console of the RSI platform. To avoid information overload, the user interface is designed to display suggestions “in a non-intrusive manner” (p. 4).

Figure 6. ASR Feature in KUDO Interpreter Assist (screenshot from Fantinuoli, 2022)



Fantinuoli et al., (2022, p. 6) in their benchmarking study using Interpreter Assist found that although the relevance of automatically extracted terms varies among evaluators, the quality of automatically translated terms remains high. The real-time suggestion feature performs well for specialized terminology and numerals, with a reported F1 value of around 98%. Nevertheless, recall values for general speeches require further enhancement. The performance for named entities increases significantly with fine-tuning. The author suggests combining automatic fine-tuning with a human-in-the-loop step to successfully integrate the tool's architecture, allowing users to add event-specific information.

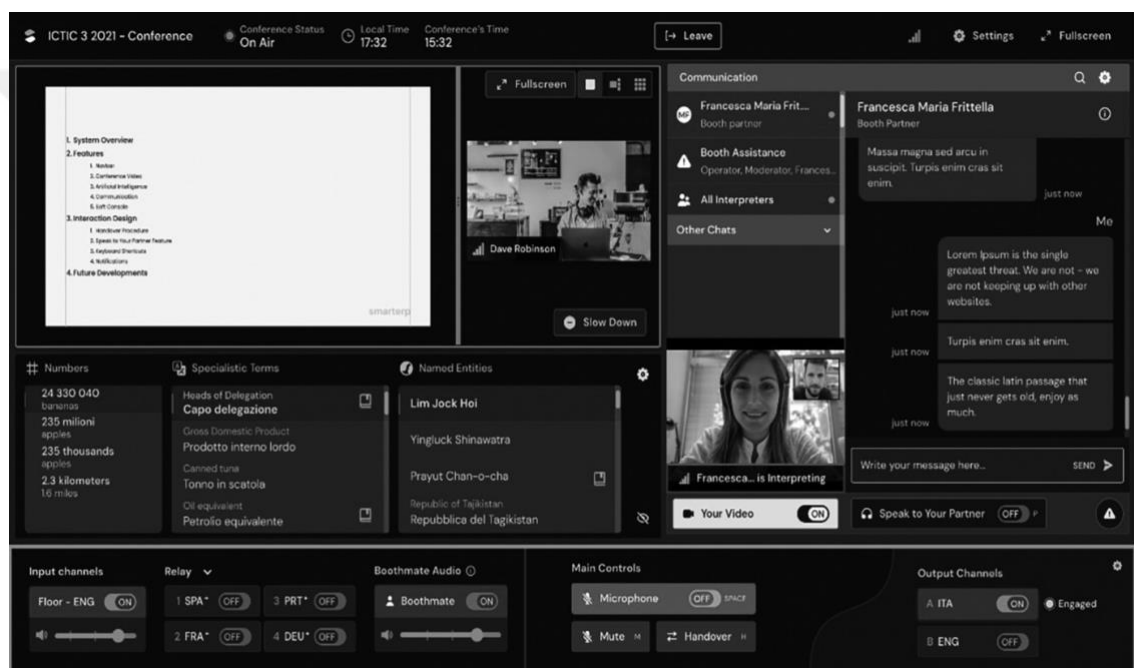
2.3.2.3. SmarTerp

SmarTerp¹⁸ is a European Union-funded project aimed at developing a remote simultaneous interpreting (RSI) platform remote simultaneous interpreting (RSI) platform that incorporates an ASR- and AI-driven CAI tool. In the scope of the project, an interdisciplinary team collaborated on the project using a user-centred design approach. The SmarTerp interface comprises two components. Firstly, the platform

¹⁸ <https://smarter-interpreting.eu>

provides a remote interpreting system with ISO-compliant audio and video quality and a communication platform (with technicians, speakers, and booth-mates) allowing interpreters to perform all necessary actions (Rodríguez et al., 2021). Secondly, the platform has an integrated ASR-enhanced CAI tool that deals with the automatic suggestion and display of named entities, specialised terms and other problem triggers (the feature commonly used in third-generation CAI tools). The project is one of the attempts to fulfil the need for integration of CAI UI and remote interpreting platform.

Figure 7. The user interface of SmarTerp



In their study, Frittella and Rodríguez (2022) evaluated the usability and user requirements of SmarTerp. The study was conducted using expert appraisal and field trial methods with eight qualified conference interpreters in a simulated RSI conference mock-up. The study presents the first evaluation study of an RSI platform integrated with a CAI tool and particularly focuses on interpreters' needs and requirements for RSI systems. The key outcomes include SmarTerp's UI features and the system's technical specifications that influenced the system's usability and participants' satisfaction with the tool. The results suggest that interpreters prioritize simplicity, naturalistic interaction with their boothmate, and the ability to operate the technological equipment strategically. The

study also highlights the need for interpreter training in the effective use of ASR/CAI tools and identifies several research gaps in the use of RSI platforms and ASR/CAI tools in real interpreting assignments. Additionally, the work of interpreters has never been analysed to identify strategic actions that interpreters need to perform when interacting with their tools through dedicated methods such as cognitive task analysis, which is a significant gap for UI design work (p. 163).

2.3.3. Speech Technologies and Automatic Speech Recognition

Speech technologies encompass a variety of tools and methods for the analysis, synthesis and recognition of spoken language. Speech technologies as an umbrella term have many ramifications, including transcription (speech-to-text), translation (speech translation) and speech synthesis (text-to-speech) services. These technologies are used in various fields, particularly in multilingual communication, and have multiple applications. A common application of speech technology is automatic speech recognition (ASR), which uses algorithms and software to convert spoken language into written text. ASR can also be thought of as a technology that enables machines to recognise and transcribe human speech. Recent research trends in computer science show a growing interest in ASR as it has been extensively studied and developed in recent years and is now widely used in a range of applications such as voice-activated assistants, dictation software and translation tools. ASR technology is also becoming increasingly accurate and able to handle a wide range of accents and languages. This has led to its use in a wide range of fields, including education, healthcare and business.

The emergence of the research and application of ASR dates back to the 50s (O'Shaughnessy, 2008). The first speech recognition system was developed by Davis et al. (1952), which could only recognise single spoken digits. As the world of technology advanced, new concepts, algorithms and techniques were created. In the 90s, a significant amount of work went into creating software that enabled research programmes around the world to produce innovative results, including the ability for machines to recognise and interpret human speech. Since 2012, ASR results flourished radically thanks to advanced deep learning which has yielded impressive results across a range of diverse

areas, including but not limited to self-driving cars and gaming, natural language processing, image recognition, and text-to-speech.

ASR technology involves a complex process of converting spoken words into written text. In the first step, the speech signal is digitized, which involves sampling the audio at a high frequency and converting it into a sequence of numbers. This digital representation of the speech is then passed through various algorithms that analyze the signal for different linguistic features such as phonemes, syllables, and words.

The key elements necessary for an Automatic Speech Recognition system to function effectively consist of three principal data sources: an acoustic model (information about the sounds of the language), a phonetic lexicon (a list of words it can recognise), and a linguistic model (an understanding of how words can be put together) (Deng and Li, 2013; Fohr et al, 2017). The former (information about the sounds of the language) includes the phonemes and other extra sounds such as pauses, breathing and background noise. The phonetic lexicon refers to the list of words that the system can recognise because there are different ways of pronouncing these words. It provides knowledge about possible pronunciations of the words spoken. The third, the language model, is an understanding of how words can be put together and help the system understand what people are saying when they use more than one word. In state-of-the-art systems, this information is learned from large amounts of data consisting of audio recordings and written texts (corpora). However, there are research focusing on eliminating reliance on human supervision, necessitating thousands of hours of transcribed speech. As such recent research has demonstrated that unsupervised ASR systems can be trained without the need for speech annotation (Liu et al., 2022).

Implementing ASR systems is a complex task, as it involves navigating the many nuances of different languages and accents. To successfully overcome these challenges, these systems require large amounts of training data to teach them the nuances of speech recognition. By analysing this data, statistical models are created that can make educated guesses about what words are being spoken based on the unique patterns and variations found in speech signals. ASR models present other inherent challenges that researchers

are continuously addressing. These challenges range from identifying distinct linguistic features, which can vary depending on factors such as speaker accent, background noise, and the speed and clarity of speech. Despite ASR systems closing in on human parity, they still struggle with difficult accents, fast-paced conversations, or highly colloquial speech. As such, ASR technology remains partly unable to capture the intricacies of human speech, such as tone, emphasis, or intent. However, one critical aspect of ASR that cannot be overlooked is the ability to accommodate language variations. Words may sound differently depending on the speaker or the context in which they are being used. To combat this issue, ASR systems utilize techniques like phonetic modelling, which associates distinctive sounds with a uniform representation. This capability enables the system to identify words, even when they are pronounced differently, ensuring that the correct transcription is made.

A noteworthy speech technology using ASR is speaker recognition, which involves the identification of individuals based on their spoken language, and natural language processing (NLP), which involves the use of algorithms and software to analyze, understand and evaluate human language. Speaker recognition has numerous applications in fields such as education, banking, healthcare, customer service, retail, and more (Wadehra et al., 2021; Singh et al., 2012).

Speech translation (ST), on the other hand, is the process of automatically transferring verbal input (speech signal) from one language into another. ST is commonly used in a wide range of scenarios such as live lecture translation (Fügen, 2008), dubbing and subtitling (Saboo and Baumann, 2019). In the past, speech translation systems were often implemented as a cascade of ASR and machine translation systems, where the output of the ASR system was fed directly into the machine translation system (Stentiford and Steer, 1988). However, this cascade architecture has been shown to have limitations, such as the propagation of errors from the ASR stage to the MT stage (Sperber and Paulik, 2020). More recent approaches to ST involve training a single end-to-end model that can perform both speech recognition and machine translation tasks simultaneously. Deep learning technology has led to the creation of a new direct ST paradigm, which involves adapting the neural networks commonly used in ASR and MT to perform ST

tasks (Bérard et al., 2016; Weiss et al., 2017). These end-to-end models are often trained on large datasets of paired speech and translation examples and can produce translations that are more fluent and accurate than those produced by cascade systems.

The proposed software in this thesis “Sight-Terp” utilizes the open-source speech translation application programming interface (API), which is one of Microsoft’s speech technology services (Microsoft Azure)¹⁹. Microsoft’s Speech Translation API ensures developers add real-time language translation into their programs through cloud-based technology. This service utilizes AI and machine learning, specifically deep and convolutional neural networks, to accurately translate the spoken language. The Microsoft speech translation API is regarded as one of the most robust ASR models equivalent, but competitor models launched by other technology giants. To name a few, Google’s ASR model is based on the DeepMind neural network architecture and is known for its high accuracy and fast processing speed. Similarly, Apple’s Siri and Amazon’s Alexa use a deep neural network (DNN) based ASR model to transcribe speech in real time. Recently, OpenAI, the company behind the famous generative artificial intelligence model ChatGPT, released a new ASR model ‘Whisper’. The Whisper model is developed to be more sensitive and accurate in dealing with accents, background noise and technical language.²⁰ Currently, OpenAI Whisper (in addition to Microsoft Azure) is one of the solutions that European Commission is using in a ongoing project for live speech-to-text translation available for greater accessibility and for automating the transcription and translation of debates into all 24 official European languages.²¹

The quality evaluation in ST is conducted mostly using automatic metrics in the field of computer science. However, there is a lack of research into the performance of simultaneous ST models in real-world communication settings. To fill this gap Fantinuoli and Prandi (2021) carried out an experiment comparing a real-time speech-to-text translation system’s output with that of human interpreters. The objective was to broaden the evaluation from automatic metrics to a more user-centric and communication-oriented

¹⁹ <https://azure.microsoft.com/en-us/products/cognitive-services/speech-translation/#overview>

²⁰ <https://openai.com/blog/whisper/>

²¹ <https://knowledge-centre-interpretation.education.ec.europa.eu/en/news/european-commission-launches-speech-recognition-project>

one. Six evaluators assessed the performance of humans and machines based on intelligibility and informativeness. The study concluded that humans outperformed the machine in terms of intelligibility, while the ST outputs performed slightly better in terms of accuracy.

Another application of speech technology is text-to-speech (TTS) commonly known as speech synthesis, which involves the use of algorithms and software to generate spoken language from written text. In the field of artificial intelligence, speech synthesis has emerged as an exceptional application that has evolved into a technology so advanced that in certain scenarios, it is almost too difficult to tell apart from the human speech it emulates (Shen et al., 2017). TTS technology has numerous applications, such as creating audio content for e-learning materials, developing of voice assistants, and assistive technologies for people with disabilities. One example might be Amazon's Polly ²²cloud-based text-to-speech service which is used for creating human-like audio versions of written content, generating voice-based customer service interactions, and creating lifelike avatars for virtual assistants. Additionally, recent advancements in cross-lingual neural codec language modelling, such as VALL-E X²³, enable the synthesis of speech in different languages while preserving the speaker's voice, emotion, and acoustic environment (Zhang et al., 2023). This innovative technology has the potential to further enhance TTS applications, opening up new possibilities for multilingual communication and voice-retentive speech-to-speech translation (machine interpreting) tasks.

TTS also constitutes the backbone of machine interpreting systems both as cascade architecture and direct speech-to-speech models. MI can be characterized as a technological advancement enabling the conversion of spoken language into another language through computer programming (speech technologies). MI involves a multi-step approach that generates an audible version of the translated text by creating a synthetic speech in the target language. Machine translation has indeed made significant improvement, but the inherent complexity of spoken discourse poses a challenge. Unlike written communication, speech is characterized by spontaneity and ambiguity, posing

²² <https://aws.amazon.com/polly/>

²³ <https://vallex-demo.github.io/>

difficulties for machines without human intervention. Furthermore, machines are incapable of inference and context anticipation, lacking the necessary background and contextual knowledge, as noted by Fantinuoli (2019, p. 342). With the ever-evolving technology, the implications of machine interpreting on the interpreting market remain uncertain. Despite the potential impact on conference interpreters' role and professionalism, the extent of the effects remains unclear, which will possibly be unveiled with new products and services that leverage LLMs and . Fantinuoli's (2019) outlook on the future of machine interpreting paints a picture of near-future where MI permeates the market from the low-end segment, where professional interpreting services are not utilized. Still, recently there are big leaps in the advancement and broad usage of MI in many settings.

On the 24th of January 2023, KUDO, a New York-based language service provider, made an announcement regarding its upcoming tool called KUDO AI, a speech-to-speech translation tool powered by artificial intelligence (AI)²⁴. The technology is designed to analyze a speaker's voice in real time, which is considered a significant advancement in the field. There is an immense amount of research being conducted on state-of-the-art speech-to-speech translation (S2ST), too, which has flourished in the booming industry of speech-to-speech translation. During the fourth quarter of 2022, the preprint research repository arXiv featured 27 papers related to S2ST (Albarino, 2023). The multitude of applications for S2ST, from real-time video call translations to AI dubbing, have contributed to the spike in research. One notable development is the release of SpeechMatrix by Meta in early November 2022. SpeechMatrix is a vast multilingual corpus of speech-to-speech translations, and its aim is to make the development of S2ST systems more accessible (Duquenne et al., 2023). With these advances, the world of language services is set to change significantly in the coming years.

2.3.3.1. ASR Integration into Translation

It should be noted that the ASR technology usage in professional tasks was not originally designed for professional translators but rather to develop and refine the system. Indeed,

²⁴ <https://kudoway.com/solutions/kudo-ai-speech-translator/>

research focused on improving the quality of ASR through different methodologies, with the professional usage of ASR largely an afterthought. Researchers investigating possible ASR usage in translation have the major focus on the impact of ASR on human translation output and its utilization possibilities on workflows, ergonomics, and productivity (Dragsted et al., 2011; Mees et al., 2013; Ciobanu, 2014; Zapata et al., 2017; Ciobanu and Secara, 2019, Ciobanu et al, 2019). According to Ciobanu's (2014) questionnaire-based study on the pros and cons of using ASR in translation services, there are a handful of professional translators who use ASR for dictating their translations into MS Word-type software (desktop-based word processors) and CAT tools. The survey in question was designed in 2014 at the University of Leeds Centre for Translation Studies and was available for over a month to be filled out by several communities of professional translators. The study concludes that the benefits of ASR outweigh the drawbacks. As such, a significant percentage of the respondents in the study reported a significant increase in productivity when using ASR, ranging from a 10% increase to a 500% increase (p. 532). In the same context, he notes that the use of the reverse method, i.e. text-to-speech can assist professionals in their revision process by reading back the translated output, which has been addressed in later years in relevant research. Relevant to this, Ciobanu in his subsequent study (2016) asserts that "ASR has the potential to enhance the productivity and creativity of the translation process, but the benefits may be overshadowed by a decline in translation quality unless thorough revision processes are implemented" (p. 124).

Research-wise, Ciobanu (2016; 2014) calls for more empirical research to investigate the latest ASR technologies and to study how professional translators incorporate ASR into their workflows and, further, claims that recent studies on productivity in this area have had design flaws that call into question the applicability of their findings. More systematic research is clearly needed for empirical validation of their choices, such as whether or not the translators focus more on the source instead of the target segments, with less dependence on translation memory and machine translation matches. Given the recent proliferation of strong speech-to-text models and commercial ASR products in the last couple of years, translation quality expectations and technological solutions may now offer different perspectives. Therefore, new studies on the impact of speech technologies

might also unveil translators' contemporary approach towards the new ASR and CAT tool integrations.

In translation dictation (TD) and post-editing (PE) context, studies provide results in favor of effective use of voice recognition. Zapata et al. (2017) conducted a pilot experiment and explored the effects of integrating MT and PE with voice recognition (VR) and TD. The study used a mixed-methods approach with a sample of native Spanish participants. The quantitative results showed that PE with the aid of a VR system was generally the most efficient method, but most participants preferred translation without the 'constraint' of MT. The study highlights issues with revision/editing times in the VR tasks, particularly due to the system's flawed transcriptions and users' lack of familiarity with TD and VR. However, the results suggest that PE with VR may be a usable way to add MT to a translation workflow, and future experiments with more participants and language pairs are planned. Overall, the study suggests that VR technology holds promise for human-aided MT and human translation environments.

Liyanapathirana and Bouillon (2022) explored the feasibility of using speech technology in the translation process for professional translators working in international organizations. Three translation methods were compared, including dictating translations with machine translation as inspiration, post-editing machine translation suggestions by typing, and post-editing machine translation suggestions using speech. The study found that using speech resulted in better BLEU scores, required fewer edits, and took less time compared to the other two methods. The study also highlighted the importance of high-quality automatic speech recognition and machine translation support for improved translation quality and productivity gains. The results provide a promising approach to integrating speech-based post-editing in translation workflows, but further research with larger sample sizes and more detailed evaluations is needed.

Recently, there are advanced studies to investigate the feasibility of speech synthesis (text-to-speech) into translation, revision, and post-editing machine translation using computer-assisted translation tools whereby the text-to-speech feature is used to read out loud the source text being translated, which aims to help the translator understand the

meaning and context of the text more accurately. Ciobanu et al. (2019), in their study, analyse the impact of sound in the source text on the revision quality, preference, and viewing behaviour of 11 participants through a case study. The researchers carried out the study with three methodological tools: error counts, a questionnaire and eye-tracking technique to see the fixation counts, and dwell time through Mean Fixation Duration (MFD), weighted average MFD, and External Resources (ER) (p. 5). The initial results of this study showed that revision quality, particularly concerning accuracy errors improved when the sound was available. The study also indicates that the use of speech synthesis seems to enhance the perception of alertness of the translators.

2.3.3.2. ASR Integration into Interpreting

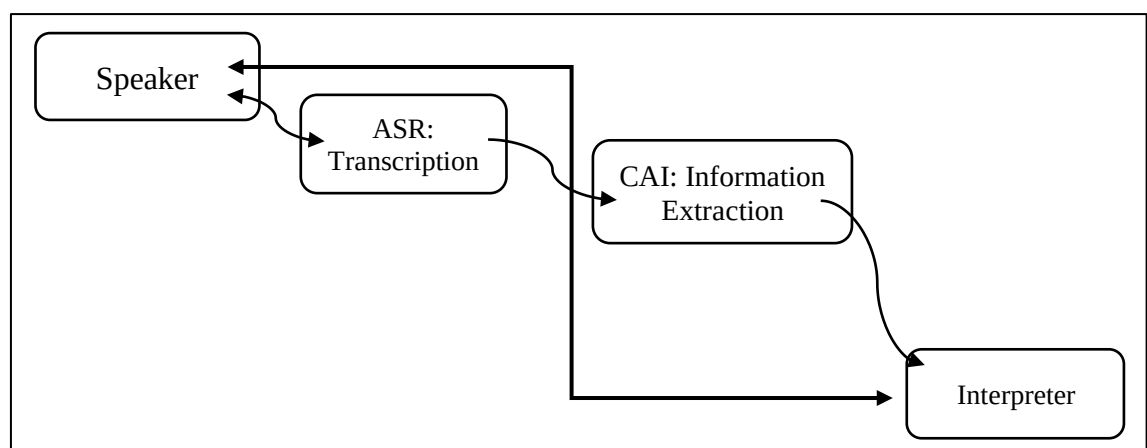
ASR in interpreting mainly stands out in two different conditions which could come as an aid: live support and preparation. In terms of live support, ASR can provide an aid which acts as, so to speak, an artificial boothmate to be used particularly in simultaneous interpreting (EABM, 2021a). In the state-of-the-art CAI tools (see section 2.3.2.), interpreters need to manually input terms or parts of them into the database to retrieve information. However, it appears to be time-consuming and distracting during an activity (simultaneous interpreting) that requires concentration and rapid information processing. An automated querying system through an ASR model has been proposed to reduce this cognitive effort (Hansen-Schirra, 2012; Fantinuoli, 2016, 2017). That is, combining AI and ASR technology can automate the query of interpreters' glossaries and provides real-time support like a boothmate and, consequently, aims to reduce cognitive effort. In addition to the *lookup mechanism*, this mechanism can also display specialised terms, acronyms, named entities and numbers, namely “problem triggers” (Gile, 2009, p. 157), for the interpreter, thereby enhancing interpreters' performance in the booth. CAI tools (InterpretBank, Interpreter Assist and SmarTerp) mentioned in sections 2.3.2.1., 2.3.2.2., and 2.3.2.3. are examples in which ASR for in-booth support is utilized with different interfaces.

Admittedly, the state-of-the-art ASR is not an infallible system and is bound to certain issues (see Section 2.4.1.1. Automatic Speech Recognition and Speech Translation).

Factors such as the type of speech (whether it is casual or formal), speaker variability, and ambiguity caused by homonymic sounds can all pose challenges for accurate transcription. Furthermore, difficulties in recognizing word boundaries can also contribute to errors in ASR output. ASR can serve different purposes depending on the constraints it has to handle. The speech at hand needs to be accurately transcribed so that the CAI tool can select pertinent text chunks for the database query algorithm and entity identification. Fantinuoli (2017) outlines the certain criteria that an ASR system needs to meet to work with the CAI tool (p. 5):

- being speaker-independent
- being able to manage continuous speech
- supporting large-vocabulary recognition
- supporting vocabulary customisation for the recognition of specialized terms
- having high-performance accuracy, i.e. a low word error rate (WER)
- being high speed, i.e. have a low real-time factor (speed of an automatic speech recognition system) (Fantinuoli, 2017, p. 5).

Figure 8. The workflow of the ASR-CAI integration in the case of InterpretBank (Fantinuoli, 2017, p. 6)



The recency of the ASR-in-process concept in SI and CI leaves much to be explored. Uncertainty of their efficacy in actual applications or how they will interact with the

intricate cognitive processes involved in interpreting (particularly SI) still remains a worth-investigating phenomenon. Research on this inquiry uses a mixed-methods approach with the question of whether the incorporation of visual prompts would result in an excessive cognitive burden, potentially impeding overall performance, or whether the reduction of other cognitive tasks by these prompts could counterbalance the added load and ultimately enhance performance. Further investigation and analysis have also been based on how training or experience on the technology can affect the utilization of such prompts.

Research on ASR support in interpreting has sparked with similar studies relevant to visual support which focused on the previously mentioned "problem triggers" (Gile, 2009, p. 157) such as acronyms, specialized terms and numbers. Numbers, among them, are notoriously taxing (Fritella, 2019, p. 81), especially in the simultaneous interpreting process. Previous studies showed that providing the numerical data as visual support during the process of interpreting can potentially decrease the error rates (Lamberger-Felber, 2001; Desmet et al., 2018). On the other hand, in terms of visual support through CAI, as stated in section 2.3.2., researchers investigated the impact of the provision the specialized terms in the booth as a visual input (manual lookup) through a product-oriented manner (Prandi, 2015; Biagini, 2015). This is particularly where ASR-enhanced CAI tools, as third-generation CAI tools (see 2.3.2) , come in handy as they can facilitate this prompting process with automatization. In the literature, it can be observed that ASR-enhanced CAI tool studies that examine the impact on the quality of the performance mostly focus on the rendition of either numbers or specialised terms.

Defranq and Fantinuoli's study explores the effectiveness of Automatic Speech Recognition (ASR) in aiding simultaneous interpreters by testing the InterpretBank ASR system, which provides real-time transcriptions with number highlighting (2021). The system was found to have high precision (96%) and low latency, meeting interpreters' ear-voice span (EVS) requirements. The study stands out as the researchers used real-life ASR system by using a commercially available CAI tool, contrary to Desmet et al. (2018) in which the ASR system is partly imitated in a mock-up scenario. The researchers examined three aspects: 1) the viability of the support offered by InterpretBank, 2)

participants' interactions with the ASR support, and 3) the effects of ASR support on participants' performance. The study involved six interpreting students and aimed to replicate a real training environment. Participants consulted the ASR output in just over half of the cases, and their interactions with the ASR support varied. The provision of ASR improved performance, with complete renditions increasing for almost all number types. The authors call for further investigation into the overall performance and participants' experiences, as the use and benefits of technological support depend on experience and expectations.

Similarly, again with a real-life ASR-enhanced CAI tool (IntepretBank ASR), Fantinuoli and Pisani in their study (2021) analyse the impact of ASR on number rendition in simultaneous interpreting with speeches dense in numbers by comparing the performance of two groups of participants, one with the aid of ASR support and the other without any technological support. An observable difference from Defranq and Fantinuoli's (2021) research is that Fantinuoli and Pisani displayed the recognized numerical output not in a text-embedded way but separately in an interface. The authors conclude a significant reduction in the error rate, which dropped from 39.8% without technological support to 14.8% with the support of ASR (p. 195). This led to a reduction in omissions, phonetic perception²⁵ errors, and a decrease in cases of approximation (p. 195). Overall, the study shows that ASR support can be effective in providing support for speeches dense in numbers and the quality of ASR-enhanced CAI output is "mature enough to be used in real-life applications" (p. 195). However, participants reported in the questionnaire that they encountered difficulties such as feeling distracted by the added visual stimulus and the need to coordinate other sub-processes lies in the interpreting process.

Research on usability and need analysis of ASR-enhanced CAI tools is another nascent field. Frittella and Rodríguez (2022) conducted a study to assess the usability and user requirements of SmarTerp (see section 2.3.2.3). The research involved eight high-level conference interpreters participating in a simulated RSI conference based on an actual European Parliament debate. This study is the first of its kind to evaluate an RSI platform

²⁵ Phonetic perception refers to the phonetically wrongful perception of a number (i.e. fifteen instead of fifty)

with an integrated CAI tool, providing insights into interpreters' needs and requirements for such systems. The key findings include the identification of SmarTerp's user interface features and technical specifications that impacted usability and participant satisfaction. The results indicate that interpreters value simplicity, naturalistic interaction with boothmates, and the strategic operation of the technological equipment.

Montecchio's Master's thesis (2021) (see Fantinuoli & Montecchio, 2022) investigated ideal and maximum acceptable latency for number-dense speech rendition with CAI tool support. In an aim to find the optimal latency in ASR-based CAI tools, the study examines its effects on rendition accuracy and delivery flow. As latency increased, both aspects suffered, indicating a higher cognitive load due to an extended ear-voice span. The study aimed to derive implications for CAI tool development and it highlights the growing interest in usability and design-focused research for CAI tools. Another example but this time focusing more on the ergonomics of CAI, The University of Ghent and the University of Mainz/Germersheim conducted an EU-funded project to develop UI recommendations for third-generation computer-assisted interpreting (CAI) tools (EABM, 2021b). They surveyed 525 conference interpreters, most of whom had over ten years of experience. The goal was to create a user-friendly artificial boothmate for interpreters. Key findings include preferences for a vertical layout with new items added below previous ones, items remaining on screen until space runs out, terms on the left and numbers on the right or both in the same box, and new items appearing in bold, larger font, or a different colour (2021b).

Using ASR for preparation is a nascent area. In this context, Gaber et al (2020) presents a new approach to using speech-to-text technology as a documentation tool for interpreters for preparation purposes. The authors offer a comparative analysis of S2T technology for interpreters and the opportunities ASR opens as a documentation aid. The study aims to establish the most suitable ASR application for building ad hoc corpora from video-recorded speeches prior to an interpretation assignment, and it introduces an approach that can lead to more future ASR-oriented interpreting research. The authors assert that ASR technology can be an effective tool for building ad hoc corpora and extracting candidate terms for interpreters.

In conclusion, ASR integration into interpreting offers promising support in live simultaneous interpreting and preparation, particularly by automating query processes, reducing cognitive effort, and enhancing performance through real-time assistance. However, challenges remain due to ASR systems' imperfections, as well as potential cognitive burdens introduced by visual prompts. Further research and development are needed to improve ASR systems and better understand their effects on interpreters' performance and experience. The growing interest in usability and design-focused research for CAI tools highlights the importance of creating user-friendly and efficient ASR-supported systems to maximize the benefits of this emerging technology in the interpreting field.

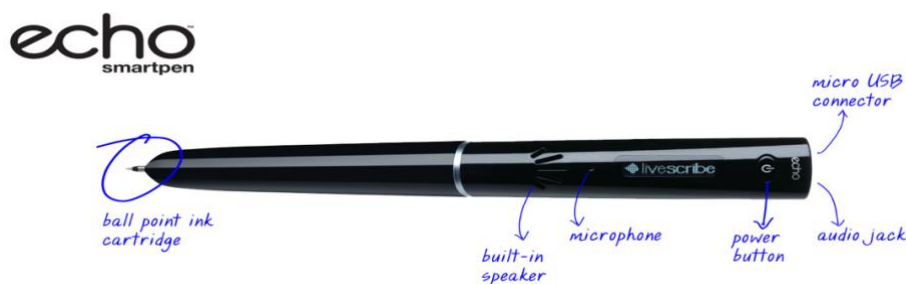
2.3.4. Technology and Consecutive Interpreting

In the literature, there is a certain number of studies scrutinizing the potential impact of technology in CI. The general research question that the studies are set to answer is whether tablets can improve the quality or what kinds of features are needed for interpreters to excel in digital note-taking (Holley & Goldsmith, 2015; Paone, 2016; Goldsmith, 2017, 2018; Arumí & Sánchez-Gijón, 2019; Altieri, 2020). In technology-mediated interpreting practices, the impact of technology seems to be more observable in simultaneous interpreting than in consecutive interpreting. The technological resources that interpreters can use for consecutive interpreting are more product-oriented since the assignment preparation of interpreters does not change depending on the mode of interpreting (consecutive or simultaneous). The nature of simultaneous interpreting in conference settings inherently would show different types of technicalities which might bring different technological support options to design, however, the technical support in consecutive interpreting is seemingly nothing more than pen and paper. Therefore, the technological support in consecutive interpreting aiding the interpreters is devised to provide a facility in either the note-taking process or the reading from notes process, which can be associated with “Listening and Comprehension phase” and “Reformulation Phase” in Effort Models (see section 2.2.) However, regardless of the difference in the fundamental dynamics in both modes, technological aid can help interpreters in their pure labour: easing cognitive load in information retrieval while increasing the quality.

2.1.4.1 Sim-Consec

The first attempts proposed a different technique namely *consecutive-simultaneous* or *simultaneous consecutive interpreting*, Sim-Consec in short. Sim-Consec refers to a mode of interpretation where a speech that would typically be rendered in consecutive mode is instead recorded, played back on headphones, and interpreted by the interpreter in real-time using the simultaneous mode. Hamidi and Pöchhacker investigated the feasibility of Sim-Consec, and in their study, three professional interpreters concluded that SimConsec “enhanced interpreting performances, reflected in more fluent delivery, closer source-target correspondence, and fewer prosodic deviations” (2007, p. 14). Additionally, related research investigated digital pen technology in this context by using a digital pen with a microphone, built-in speaker, recorder, infrared camera, and ink (Hiebl, 2011; Orlando, 2014; Mielcarek, 2017). While the normal note-taking process is underway, the audio data coming from the speaker is captured thanks to an embedded chip. Once the notes are taken, the data and notes become synchronized. The interpreter can then play back the audio recording while simultaneously reviewing the notes. In an empirical study by Orlando (2014), the Smartpen Livescribe™ Echo® is used in Sim-Consec assignments delivered by four professional interpreters. The study concludes that this new mode can help interpreters to render more accurately, comprehensively, and fluently. Orlando observes greater accuracy, and fewer disfluencies or hesitation phenomena, as well as increased interpreter confidence and a more complete rendition of the interpreted material (2014). Nonetheless, there is still a need for more empirical studies on the effectiveness of the digital pen, as well as the inclusion of a greater sample size.

Figure 9. The functionalities of Livescribe™ Echo® Smartpen



3.1.4.1 Tablet Interpreting

Academic studies on technology use in consecutive interpreting have recently emerged. Therefore, there are limited studies on tablet interpreting apart from some studies (e.g., Rosado, 2013; Goldsmith and Holley 2015; Paone 2016; Ocegüera López 2017; Goldsmith 2017; Dreschel & Goldsmith, 2016). In their exploratory and mixed-methods study (2015), Goldsmith and Holley designed a set of features to use while evaluating the applications, styluses, and tablets for consecutive interpreting. They reported that the tablet can be used in several contexts, and it can outstrip the pen-paper method in terms of functionality. The study enquired how tablets can comply with note-taking requirements, how professional interpreters use tablets, and why. The study reports that tablets “can equal pen-and-paper interpreting in many contexts and settings” (Goldsmith and Holley, 2015; Goldsmith, 2018, p. 360), especially with newly added functionalities. As a result of the study, Goldsmith and Holley draft a concise compilation of the advantages and drawbacks of tablet interpreting (2015).

Table 2. *Advantages and Disadvantages of Tablet Interpreting (Goldsmith, 2018, p.357)*

Technical advantages	Technical disadvantages
Button with extra features in stylus Cloud backup and accessibility Cut and paste Email and print notes Eraser Internet connectivity Password protection Split-screen Unlimited ink and paper	Crashes Inadvertently hit menu buttons Multiple cables needed Need to prepare equipment Embed notes
Visual advantages	Visual disadvantages
Vertical scrolling Multiple ink colours + pen styles in one stylus Hard to see in some environmental conditions Variety of paper types Custom paper Zoom (increase writing size) Zoom (see multiple pages at once)	Difficult to find place when scrolling Inaccurate/imprecise writing Inaccurate/imprecise writing Stray marks
Physical disadvantages	Physical disadvantages
Less cumbersome	Cumbersome when standing or moving

Lightweight and small	Different writing position
Client relations – advantages	Client relations – disadvantages
Impresses clients Sets interpreter aside as memorable Looks more professional Less noticeable page turns Quieter than paper	Confidentiality concerns Clients mistrust/unfamiliar with technology Clients fear interpreter is unfocused/cheating
Other advantages	Other disadvantages
Environmentally friendly Facilitates preparation Storage and organization of notes	Additional stress (fear of crashing/breaking) Cost Lack of training courses Learning curve Possible distraction

Goldsmith, in the second phase of his MA Thesis presents the first comparative user evaluation of tools used by interpreters using tablets in consecutive interpreting (2017). The survey, designed to examine the landscape of tablet users in consecutive interpreting, shows that interpreters prefer note-taking applications that are reliable, durable, and comfortable (p. 49). Results also indicate that the interpreters most frequently use iPad pro and the stylus Apple Pen. GoodNotes, Notability, Notesshelf, and Penultimate are the note-taking applications that the participants scored more. Nebo, the base note-taking application used by Sight-Terp, is one of the applications scored in the study in terms of functions. A very recent example study by Mirko Altieri (2020) has shown that participants using pen and paper performed slightly better in consecutive interpreting than with tablets. Future research involving more subjects, other pairs of languages and tools might glean more data on the feasibility of the tablets. On the other hand, Arumí and Sánchez-Gijón (2019) concluded in their pilot study that participants' digital tablet usage did not negatively affect the clarity of the speech structure. Moreover, in the study, most of the participants in group A reported that the possibility of changing the thickness and the colour of the pen is an advantage for note-taking. In their 2016 paper, Drechsel and Goldsmith (p. 17) propose an intriguing hypothesis that the use of a 'streamlined device' could potentially reduce cognitive load by allowing interpreters to focus on the essential

aspects of the job. However, the authors acknowledge that further research is needed to test the validity of this proposition.

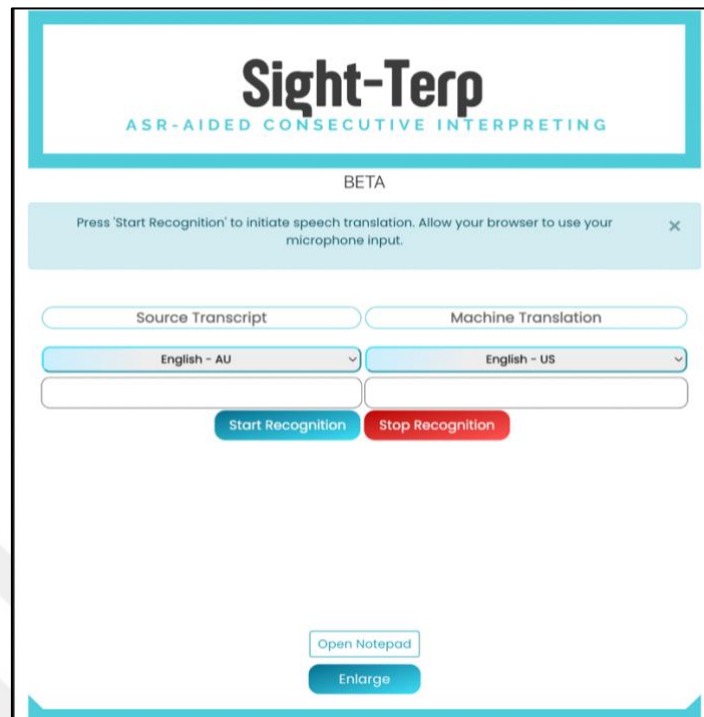
Digital note taking has also become a concept which is investigated from training perspectives. As such, there are studies conducted to propose and address the implementation of the digital pen in consecutive interpreting classes, especially when it is used in connection with the development of note-taking techniques (Orlando, 2015a, 2015b; Ocegüera López, 2017).

2.4. SIGHT-TERP

The proposed software Sight-Terp is a prototype of a web-based ASR-enhanced computer-assisted interpreting tool designed and developed by the author of this thesis. Sight-Terp initiates continuous speech recognition and transcribes the source speech input by the speaker and automatically generates a machine-translated output of speech segments, creating two automatically-generated reference texts.

The tool can be used on tablets, mobile phones, and computers and is designed to be used in interpreting scenarios where consecutive interpreting modality is used. The Figure 9 details the main layout of the page and the buttons which are bound with certain additional functions. The “Start Recognition” button commences the speech translation session. Two adjacent text boxes are positioned to display ASR and speech translation (using neural machine translation) results. While the recognition is running, the named entities found in the text are highlighted in real-time through named entity recognition. On the bottom, an optional, third-party digital note-taking application is embedded which, if preferred, can be incorporated into the program's functionality. It has certain functions such as erasing with a scratch-out and drawing lines by underlining and circling. The digital notepad is intended to give the same feeling and experience as conventional pen-paper in consecutive interpreting. A tablet with a stylus allows for quick and flexible handwriting on Sight-Terp and swift and adaptable handwritten annotations, providing substantial assistance to interpreters in consecutive interpreting. In this section, the overall design of Sight-Terp and its general features will be described with the rationale behind each object.

Figure 10. The main layout of Sight-Terp (Tablet View)



2.4.1. General Features

The CAI tool Sight-Terp, in its minimalistic design, has three main features. The first is speech translation through ASR along with text segmentation, the second is called named entity recognition (NER and highlighting), and the third is a digital notepad. Speech translation, the main feature of the tool, creates an automatic reference text in both source and target languages to provide a reference text aid that aims at improving interpreting performance. The reference texts are displayed in segments, making them easy to be read and processed by machine translation. The named entity recognition (NER) feature recognizes and highlights the named entities in the text. Highlighting is done to help the user to detect critical content information such as organizations, proper names, and numbers. The digital notepad is an optional feature that creates an area and allows the user to draw or take notes using a stylus. All the features are detailed and elucidated in the subsections below.

2.4.1.1. Automatic Speech Recognition and Speech Translation

Automatic speech-to-text translation, also known as speech translation (ST) (see Chapter 2.3.2.), constitutes the main function of Sight-terp. As shown in Figure 10 the interface contains two text areas side-by-side, source transcript and machine translation. When the start recognition button is pressed, it initiates an automatic speech recognition session using Microsoft Speech Translation API²⁶, which is based on an end-to-end system using deep neural network-based modelling. The direct speech translation method using such end-to-end trainable encoder-decoder models is reported to outperform the conventional cascade approach where ASR and MT are used separately in a non-unified way (Bérard et al., 2016; Weiss et al., 2017).

The automatic speech translation model integrated with the tool conducts real-time multilingual translation through continuous speech recognition. The speech recognition model is JavaScript-based so it can be used through browsers. The API uses the default system microphone for audio input. When the input voice is received, the browser sends a request to the server through WebSocket protocol to reach Microsoft servers located in Europe. As soon as the communication is secured with one persistent connection, the speech translation starts. Firstly, the model tries to recognize a single utterance with minimal latency²⁷. A single utterance might contain up to three or four sentences or only one. When the speaker gives a full pause or slight silence, the model recognizes this as the end of the utterance. This continuous speech recognition keeps on predicting whether the sentence has come to an end as it continues to listen to the auditory input. This prediction is made possible using “a learning adaptive segmentation policy” (Zhang et al., 2022) where translation starts when a meaning unit is detected, in contrast to fixed policy (Nguyen et al., 2021) where speech utterances are split at a fixed frequency.

When the speech units are recognized to be full and semantically meaningful, the AI-based speech model automatically punctuates the sentences, thus making them meaningful and complete. Once completed, the same process starts for the other

²⁶ <https://learn.microsoft.com/en-us/azure/cognitive-services/speech-service/how-to-translate-speech>

²⁷ Latency might be more based on the internet connection health. Unsolid broadband connections might affect the transcription/translation results badly. Generally, it is advised to have a strong internet connection while using Sight-terp.

consecutive speech segments. In other words, the speech tokens are sequenced to make up speech segments forming an utterance and presenting it a complete sentence at the backend. These speech recognition session outputs are concurrently translated through neural machine translation without any delay.

Finally, the speech recognition results (source and MT text) concurrently go through a text-normalization²⁸ and automatic punctuation processes before appearing as the final phonemic and/or textual representations of the source input. This is made by Microsoft's TrueText function which adds sentence breaks and removes disfluencies ("Uhm"s and "uh"s)²⁹ and stutters in the raw ASR output. Automatic punctuation as well as automatic capitalization prediction are important for other natural language processing tasks like machine translation and should be based on acoustic information such as pauses and pitches to prevent ASR and segmentation errors (Nozaki et al., 2022, p. 1). The speech translation API used in Sight-Terp successfully recognizes the most-used punctuations like commas, interrogation marks, and dots in accordance with the intonation and pacing of the source input. By that means, punctuation enhances the readability of the text by humans and therefore increases the accuracy rate of neural machine translation by eliminating potential contextual errors. In addition, thanks to correct capitalization both in sentence starts and proper names, the named entity recognition feature (see Named Entity Recognition and Highlighting below) works in the best possible way.

The source transcript and machine-translated output that is generated through direct speech translation are displayed on the main interface in real time, thus creating two reference texts for the user (interpreter). This feature of Sight-Terp intends to provide additional reference text(s) for the interpreter in the consecutive. Thus, the tool intends to

²⁸ Text normalization, or standardization, refers to the process of mapping non-standard words, such as symbols, numbers, and abbreviations, to standard words that are pronounced in a consistent manner through strings of characters. (e.g., "The budget is \$500 and we will buy 2lbs of bananas" is converted to "The budget is five hundred dollars and we will buy two pounds of bananas.")

²⁹ Although certain disfluencies are omitted automatically in the speech recognition between the sentences, very long and repetitive disfluencies within the sentences such as 'uhm's and 'uh's constitute a problem as the ASR does not guarantee full accuracy at predicting whether the sentence ends or not. This problem inherently influences the machine translation output as well. This is why the recognition accuracy can be expected to be less in informal speech.

help the user to render the source text in a *sight-consecutive* modality by improving the lookup mechanism and providing a memory prompt in both source and target languages, especially in long consecutive interpreting.

2.4.1.2. Automatic Text Segmentation

Automatic text segmentation allows both source and machine translation texts to be displayed concurrently in a vertical form in the adjacent text boxes of Sight-Terp during continuous speech recognition. That is as the speech unfolds into complete texts, the final text-based outputs are simultaneously displayed in segments with ordinal numbers starting from number one. Each segment is formed in each big silence (approximately 2 seconds). This feature is deployed with the aim to display the reference text in an easy-to-read fashion and allow the user (interpreter) to follow up the source segment with its target MT output thanks to the enumerated style.

Figure 11. A segmented text on the interface of Sight-Terp

Source Transcript	Machine Translation
English - US	Turkish - TR
1) Vegetarianism is the practice of not eating meat or fish. People who follow vegetarianism are called vegetarians. Vegetarians eat foods like vegetables, fruits, nuts, beans and grains. 2) There are many reasons for not eating meat. Some people think that it's wrong to kill animals, other think that eating meat is bad for their health or the world. 3) This is because the land used for animals can be used to grow food. 4) Some people might not become vegetarian because their religion says not to eat animals. Vegetarians who do not drink milk or eat eggs are called vegans. 5) Vegans also often will not use animal products like leather, but many vegetarians use animal products.	1) Vejetaryenlik, et veya balık yememe uygulamasıdır. Vejeteryanlığı takip eden insanlara vejetaryen denir. Vejetaryenler sebze, meyve, fındık, fasulye ve tahıl gibi yiyecekleri yerler. 2) Et yememenin birçok nedeni vardır. Bazı insanlar hayvanları öldürmenin yanlış olduğunu düşünürken, diğerleri et yemenin sağlıkları veya dünya için kötü olduğunu düşünüyor. 3) Bunun nedeni, hayvanlar için kullanılan toprağın yiyecek yetiştirmek için kullanılabilmesidir. 4) Bazı insanlar vejetaryen olmayabilir çünkü dinleri hayvanları yememeyi söyler. Süt içmeyen veya yumurta yemeyen vejetaryenlere vegan denir. 5) Veganlar ayrıca genellikle deri gibi hayvansal ürünler kullanmaz, ancak birçok vejetaryen hayvansal ürünler kullanır.
Start Recognition	Stop Recognition

The similar vertical segmentation approach is generally accepted in conventional note-taking process as suggested by Jean Herbert (1952) in his book “Interpreter's manual: How to become a conference interpreter”³⁰ and Jean-François Rozan (1956) in his book titled “Note-taking in consecutive interpreting”³¹. Depicting the logic and the idea in a vertical fashion (from top to bottom rather than left to right) on the page using “shifts” (Rozan, 1956) creates a structured note and helps jog the interpreter’s memory better. Using horizontal lines segmenting the ideas or the sentences from one another, on the other hand, helps the interpreter analyse the order of the ideas easily in her/his mind. Indentation, spacing, and vertical organization in the notepad during the note-taking process are generally accepted by many tutors and scholars. Verticality and sectioning with lines are suggested by other scholars such as Roderick Jones (2002), Dörte Andres (2002) Christopher Thiéry (1981). Andres, for example, argues that “The segmentation and the arrangement of the notes on the page can facilitate assignation (of the meaning) and have a positive effect on oral reproduction” (Andres, 2002, as cited by Gilles, 2017, p. 277). The auto-segmented outputs on Sight-Terp are not displayed as large chunks but as segments making the user interface (UI) organized as a whole. Sight-terp, in the same manner, is designed for the same facility: segmentation in the reference text aid helps the interpreter to read and grasp the needed information and skim through the speaker’s line of argument thoroughly.

Rozan states that “The first rule of consecutive interpreting is that the real work must already have been done when you start reading back your notes: the text, its meaning and the links within it, must have been perfectly understood.” (1956, p. 27). It is advised that the interpreters in the consecutive mode note the ideas making the use of the “meaning units” (Seleskovitch, 1989) and they also note the non-contextualized information like figures and proper names to help their memory in the event of oral reproduction. In Sight-Terp, the full source text and the machine-translated output of the speech are presented in the adjacent boxes on the screen. As both automatically generated reference texts are displayed in enumerated fashion, it is also possible to find the target text equivalence of the units of interest.

³⁰ In its original name: Manuel de l’interprète: Comment on devient interprète de conférences

³¹ In its original name: La prise de notes en interprétation consécutive

In their study, Xinyu Wang and Caiwen Wang (2018) investigated whether a possible MT reference in consecutive interpreting might boost interpreting accuracy. Participants were provided with the reference machine translation text as a full paragraph using ASR beforehand. In the post-experiment questionnaire, 9 out of 10 participants reportedly failed to locate needed information which resulted in hesitations and pauses and, eventually, lower fluency scores (Wang & Wang, 2018). This data is a clear indicator of the difficulty to carry out interpreting with an unsegmented long paragraph. As recommended by the author and the participants in the study, texts displayed in chunks by sentence or utterance would be better for the look-up mechanism (p. 136) as well as facilitation in focusing on each translatable item.

As mentioned earlier, the segmented chunks enable the user to relocate the corresponding MT text by following the numbers of the segments. On top of that, when encountered with a proper name (e.g., name of an organization) or a field-specific term in the source speech, the interpreter can benefit from the corresponding enumerated MT output without losing the textual integrity of the speech and read from the screen. This feature also allows to organization of the speech (if suitable) vertically in the source text box and helps the interpreter to produce the target text without getting lost in the ASR-outputted full text, which was observable in Wang & Wang's study.

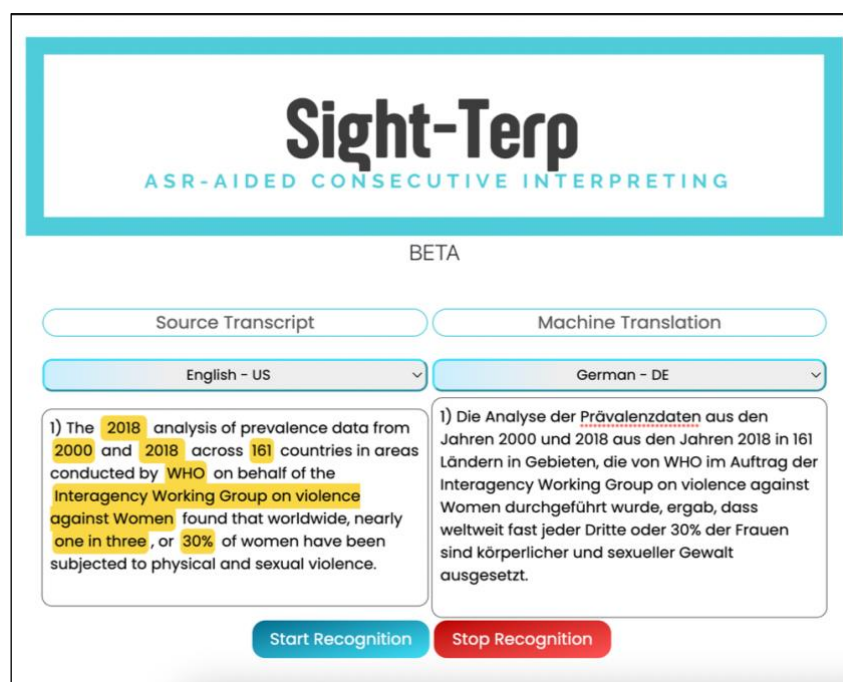
Admittedly, an overreliance on the segmented text with its machine translation in coping with two references might cause a cognitive load in the interpreter's coordination and production effort (see section 2.2.), which future experimental research with Sight-Terp might unveil. This potential drawback might bring about more cognitive load and therefore less accuracy. As a matter of fact, the results of this experimental study also will indicate whether higher information input in computer-assisted consecutive interpreting (with references) results in less accuracy in the interpreters' performances.

2.4.1.3. Named Entity Recognition and Highlighting

Named entity recognition is a computational sub-task used for information extraction from a raw text, which basically identifies the pre-defined entity categories from the text

spans (Kim Sang et al., 2003; Cui et al., 2021). Sight-Terp uses a NER model from Microsoft Cognitive Services Text Analytics API³², which is a neural NER model giving high confidence scores in generic texts. Such convolutional neural network-based predictive models require large, labelled training data. The NER model implemented in the speech translation module of Sight-Terp recognizes the entities in the unstructured text based on certain categorizations such as places, people, organizations, and numbers. The process is as follows: right after each speech segment is displayed in the result boxes, the chunk (source) text is sent to A Node.js application running on a different server. This server-side JavaScript application recognizes the entities in the raw text and in return creates a heterogeneous array of results. Finally, the main application listens to the server through the WebSocket communication protocol. If there are entities from the pre-defined categories, the recognized entities are directly highlighted in Sight-Terp's main interface while the speech recognition is still under-way for the subsequent speech segments. The categories that the model seeks to highlight in the text are organization names, person names, dates, numerical data (e.g., percentage, ordinal numbers, temperature), location names, and currency data (e.g. two million \$).

Figure 12. Named entities highlighted in Sight-Terp interface



³²<https://learn.microsoft.com/en-us/azure/cognitive-services/language-service/named-entity-recognition/overview>

The highlighted entities can also provide a backup reference text for cases when the full ASR results are not resorted. Highlighting is designed to facilitate the interpreter's task in the production effort as the task is mainly done on the sight interpreting modality, where there needs extra effort while reading from the script and at the same time rendering it orally. Accordingly, the categories to be recognized by the model (organization names, person names, proper names, dates, numerical data etc.) are the units that have critical technical and contextual information. These units are also the units of interest. The highlighting feature is therefore implemented to ease this 'reading from notes' effort by facilitating the reformulation of the message while reading, which can be unveiled with different types of studies using Sight-Terp.

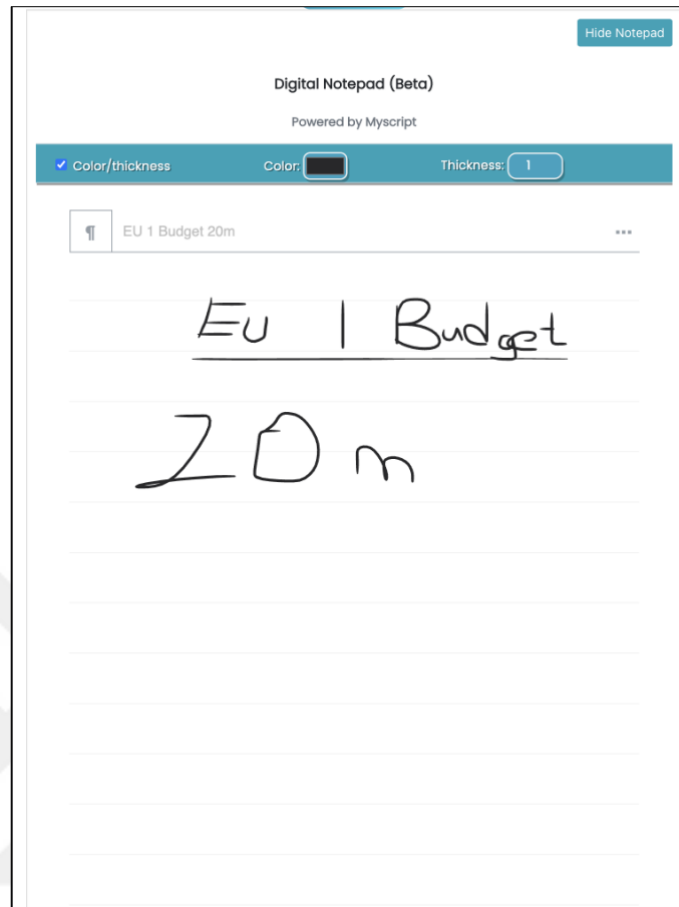
2.4.1.4. Digital Notepad

The digital notepad in Sight-Terp is positioned below the main layout. The web application of the note pad is based on iinkJS³³ JavaScript library provided by Myscript³⁴, which runs fully on cloud in a client-server configuration and has certain functions like handwriting recognition, digital ink capture and rendering with 65 supported languages. The only way to use the notepad effectively is using a stylus like Samsung S pen or Apple Pen, which give reality-based handwriting experience. As a matter of fact, digital notepad would be more effective and suitable if used in tablets.

³³ iinkJS is a JavaScript (programming language) library used primarily for handwriting recognition.

³⁴ <https://myscript.github.io/iinkJS/docs/>

Figure 13. Digital Notepad feature of Sight-Terp



Myscript's features for math is used for arithmetic calculations and other math expressions presented for different users, which, in a sense, are out of the scope of the type of note taking that is used in consecutive interpreting. Rather, the tool is implemented for the features for text writing which includes flexible text entry and automatic suggestion through recognition engine. Among all features, the digital notepad, which is customized for Sight-Terp, is designed only for note-taking action during interpreting assignments. That is, only digital ink, automatic text recognition, and gestures are implemented in the application. Gestures in this context mean pen actions or sets of strokes to edit or decorate content like crossing off or striking out. For example, scratching out a text erases the text block, drawing a frame around a word or underlining it highlights the text. It is known that the pen actions aforementioned are used by interpreters quite often while taking notes, either to emphasize the meaning or indicate that the content has a negative/positive meaning. By clicking on the three dot shaped icon,

the user can clear the page by pressing the button “Clear”. This can be repeated in each turn-taking in consecutive interpreting.

It is worth mentioning that the experiment procedure of this study does not include any test on the digital notepad, as the main research questions of this thesis concern usability of ASR and NLP applications without involving basic note-taking with a pen. However, the notepad is implemented to allow users to be able to take notes while using other ASR features at the same time. In other cases, Sight-Terp can also function as a sole digital note-taking application. As the ASR engine is not infallible, note-taking in necessary settings is still needed. Inarguably, future empirical studies with Sight-Terp might offer explorations on the interoperability of the digital notepad with a speech translation module. Although digital note-taking is discarded for now, the speech technologies within the scope of this study will be tested not using laptops but tablets. Therefore, the answers from the questionnaire in this study on the ergonomics of using a tablet in a consecutive interpreting assignment can yield illuminating results which can be compared to the results of the academic studies conducted on this particular phenomenon. The literature review on the usability of tablets and interpreters’ preference on effective features are highlighted in section 2.3.4.

CHAPTER 3

METHODOLOGY

3.1. DESIGN OF THE STUDY

The present study employed an experimental design to investigate the impact of an ASR-enhanced computer-assisted interpreting tool on the accuracy and fluency of sight-consecutive interpreting performance of novice interpreters ($n=12$). The study is conducted in the repeated measures design, with the pre-test serving as a baseline for comparison with the post-test scores. The stimuli used for the interpreting tasks were pre-recorded speeches in the direction of English into Turkish, which were validated for their level of difficulty through various readability indexes. The experimental design aimed to evaluate the effectiveness of the Sight-Terp tool (see section 2.4. Sight-Terp) in enhancing the interpreting performance of the participants while minimizing any potential confounding variables.

In terms of investigating the impact of the CAI tool aid on the accuracy of interpretation, this study aims at comparing the accuracy of interpretation between two conditions (with and without technology support) for the same set of participants. The within-subjects factor is the condition (with or without technology support), and the dependent variable is the accuracy of interpretation, measured as the percentage of accurately rendered units of meaning, and fluency (see 3.5. Data Analysis). As a result, the independent variable is the technology usage (the ASR-enhanced CAI tool Sight-Terp on a tablet), while the dependent variable is the interpreting performance of the participants. In the test, each participant was asked to interpret a speech without any technological aid (pre-test), and then receive training on how to use the Sight-Terp tool. After the training, the participants were asked to interpret another speech on the same topic using the Sight-Terp tool (post-test).

Given the nature of our data and the specific requirements of the research questions, non-parametric statistical method is applied, which do not assume a normal distribution of data. It appears that the data for the Sight-Terp conditions using both Speech A1 and Speech B1 are normally distributed (since the p-values for both the Kolmogorov-Smirnov and Shapiro-Wilk tests are greater than 0.05). However, the data column for the No Tech. Aid conditions using both Speech A2 and Speech B2 do not appear to be normally distributed (since the p-values for both the Kolmogorov-Smirnov and Shapiro-Wilk tests are less than 0.05). Since not all of the data are normally distributed, using a non-parametric test would be more appropriate. Additionally, the Wilcoxon Signed-Rank Test was chosen as the most appropriate statistical test due to our paired data design. A within subject statistical test is appropriate for this study because it allows for the examination of the effect of the condition on the accuracy of interpretation while controlling for individual differences among participants. By using the same set of participants for both conditions, it is possible to minimize the influence of individual differences, such as language command, familiarity with technology in general or experience in interpreting, on the accuracy of interpretation.

The test will be conducted two times in a row with different but similar speeches each time. Conducting two pre-post test designs consecutively with different but similar stimuli can be useful in a few ways. Firstly, it can help increase the reliability of the results by reducing the effect of random variability or measurement error that might occur in a single pre-post test design. By repeating the experiment with the same participants, it will also be possible to compare the results of the first and second pre-post-test designs to check if they are consistent. Secondly, it can help to identify the learning effect or any other factors that might influence the results over time. The study hypothesises that the use of the ASR-enhanced CAI tool in consecutive interpreting improves the accuracy of interpretations with a loss in the fluency of rendition.

3.2. DATA COLLECTION INSTRUMENTS

The data collection instruments for this study will consist of three main instruments: the CAI tool Sight-Terp, speech materials as stimuli and a rating scale questionnaire. The Sight-Terp tool were used during the post-test to assist the interpreters in their

interpretation phase by automatically creating two reference texts in a segmented and aligned style and visually displaying the named entities in the speech. Speech materials would refer to the actual audio recordings or transcripts of the consecutive interpreting tasks that were performed by the participants, both with and without the aid of the Sight-Terp tool. The performance of the participants using Sight-Terp will be recorded and analysed to assess its effectiveness in improving the accuracy and fluency of consecutive interpretation tasks. In addition, a rating scale questionnaire will be used to collect subjective feedback from the interpreters on their experience using the Sight-Terp. The questionnaire consists of Likert scale questions, as well as open-ended questions, allowing the interpreters to provide detailed feedback on the tool's usability, reliability, and effectiveness. The collected data from the two instruments are triangulated to provide a comprehensive evaluation of the effectiveness of the Sight-Terp tool in improving the performance of consecutive interpreters.

3.2.1. Speeches

The materials used in the experiment consist of four speeches to be delivered in English by a native speaker and interpreted in the consecutive interpreting modality into Turkish, which is the participants' mother tongue. The speeches are classified under two broad subject titles, with two separate speeches on each topic. The first two speeches address the issue of violence against women, while the third and fourth speeches focus on earthquakes in Japan³⁵. These speeches provide two topics with a diverse range of content to be interpreted, allowing for a thorough evaluation of the consecutive interpreting process.

The consistent level of difficulty across the speeches is crucial for ensuring a fair evaluation of the interpreters' performance. To ensure the validity of the materials, various readability indexes were applied to all speeches. Despite this, the results generally show closely comparable ratios, indicating that the speeches are at a consistent level of difficulty. The Automated Readability Index (ARI) is a computational tool used to assess the readability of written text. This index considers several factors, including the average

³⁵ The transcriptions of the speeches are in the Appendix 1 .

number of characters per word, the average number of words per sentence, and the grade level of the text, to determine its readability score. Conversely, the SMOG index, (acronym for Simple Measure of Gobbledygook), assesses the readability of text by evaluating the number of polysyllabic words present in a sample of text. In addition, the Flesch-Kincaid Grade Level formula considers the average syllables per word and the average words per sentence to assess the readability of a given text. Meanwhile, the Coleman-Liau index measures the readability of text by assessing the average number of characters per word and the average number of sentences per paragraph. In contrast, the Gunning-Fog Index calculates the readability of text based on the average number of words per sentence and the percentage of complex words in the text. Finally, the Flesch Reading Ease formula assesses the readability of text based on the average number of syllables per word and the average number of words per sentence, giving a score between 0 and 100, with higher scores indicating easier readability. In addition to the readability index, lexical density levels of each speech were calculated and compared to ensure that all speeches have moderate and equal lexical density ratios. This is done to reduce the influence of other factors on the dependent variable and increase the likelihood that any differences in the results can be attributed to the independent variable. A careful selection of materials and keeping the difficulty close to each other ensures internal validity. Thereby, it is also to secure the consistency and stability of the research methods and results by preventing confounding variables that could affect the results. All readability index results as well as lexical density results are listed in the Table 3 below.

Table 3. *Readability Index Results and Lexical Density Ratios of Speech Materials*

Reading Index	Subject: Earthquakes in Japan		Subject: Violence against Women	
	Speech A1	Speech A2	Speech B1	Speech B2
Automated Readability Index	9.47	10.75	9.06	9.56
SMOG	10.91	11.13	11.15	11.71
Flesch–Kincaid Grade Level	8.88	9.24	8.5	9.66
Coleman-Liau Index	10.61	12.11	11,08	12.46

Gunning-Fog Index	11.12	11.40	11.24	12.14
Average Grade Level	10.2	10.93	10.21	11.67
Median Grade Level	10.61	11.12	11.08	12.06
Flesch Reading Ease	60.207	58.298	56.084	40.906
Lexical Density	51.57%	56.09%	50.00%	54.93%

The speeches have similar durations and contain an indefinite but slightly equal amount of named entities and numerical data, depending on the content of the subject. This means that the interpreters will need to accurately convey the named entities and the numerical data included in the speeches, which, in a sense, adds an additional level of challenge to the interpretation task. It also gives an opportunity to answer another possible research question as we can see if there are any differences between the test performances of the group of participants in terms of the number of interpreted named entities such as location names, person names, numerical data, and organization names. The inclusion of these elements in the speeches is also important for ensuring the accuracy and completeness of the interpreted message and for conducting a thorough analysis of the accurately interpreted/rendered units of meaning in the post-evaluation process. The characteristics of the materials used are given in Table 4.

Table 4. Detailed Descriptions of Speech Materials (Duration, Length, Units of Meaning)

Material Name	Duration	Length (Words)	Number of Units of Meaning
Speech A1 Earthquakes in Japan	04:29	465	109
Speech A2 Earthquakes in Japan	04:35	452	127
Speech B1 Violence against Women	04:01	513	159
Speech B2 Violence against Women	03:39	404	125

Before conducting a study involving ASR, it is important to evaluate the accuracy of the system by calculating the word-error-rate (WER) for each speech. The WER calculates the percentage of incorrectly recognized words compared to the total number of words in the speech. This metric can provide an objective measure of the quality of the ASR system and identify potential issues that may affect the study's results. Additionally, calculating the WER can help ensure the difficulty levels of the speeches are similar. If the WER for one speech is significantly higher than the other, it could indicate that the speech is more challenging to transcribe accurately using the ASR system. In this case, it is essential to consider adjusting the difficulty levels or taking other measures to ensure equivalence. Table 5 shows the WER rates for the speeches used for the post-test where the tool Sight-Terp is used. The table also indicates the precision of ASR on the named entities (proper names, numbers, acronyms) which are the essential parts of the speech and interpretations.

Table 5. *Word-Error-Rate Results and Precision of ASR in Named Entity Recognition*

Material Name	Word-Error-Rate (WER) by ASR	Named Entity Precision by ASR
Speech A1 Earthquakes in Japan	N/A	N/A
Speech A2 Earthquakes in Japan	9.7%	30/30
Speech B1 Violence against Women	N/A	N/A
Speech B2 Violence against Women	7.4%	30/32

3.2.2. Questionnaires

The questionnaire comprises Likert scale questions and open-ended questions to gather comprehensive feedback on the tool's effectiveness, usability, and reliability. The survey administered to participants is applied in order to uncover any potential factors that influence or challenge their performance or learning processes. Incorporating expert opinions into the development of our questionnaire was done to ensure its validity and relevance. The questions are as follows:

1. How would you evaluate your experience with Sight-Terp?
2. (Likert) I think that Sight-Terp is an easy-to-use tool.
3. (Likert) Using automatic speech recognition (ASR) during consecutive interpreting tasks negatively impacted my performance.
4. (Likert) I believe that the functions available on Sight-Terp contributed to my consecutive interpreting performance.
5. Do you think that the automatic speech recognition (ASR) function in Sight-Terp is accurate and reliable?
6. Which automatically generated output did you use for support during consecutive interpreting?
7. Would you use the Sight-Terp tool in your future professional life?
8. Is there any feature/function that you would like to see in Sight-Terp?

3.3. PARTICIPANTS

Twelve participants were recruited for this study using convenience sampling. All participants were third and fourth-grade students in the English Translation and Interpreting (TIS) program. All participants were recruited from Istanbul Yeni Yüzyıl University and only students who had achieved BB grades or higher in the "Introduction to Consecutive Interpreting" and/or "Note Taking for Interpreting" course were chosen as participants. Participants who did not meet the aforementioned criteria were excluded from the study. By focusing on this specific subset of students, the study aimed to better assess the impact of Sight-Terp on overall performance of participants with note-taking and consecutive interpreting techniques. Of the 12 participants, 5 are female and 7 are male. The participants' ages ranged from 20 to 24 years old, with a mean age of 22 years old. Table 6 shows the distribution of the speeches in two conditions per participants.

Table 6. *Distribution of Speech Materials per Participant*

Participants	Earthquakes in Japan		Violence against Women	
	Speech A1	Speech A2	Speech B1	Speech B2
Interpreter 1	No Support	CAI Support	No Support	CAI Support
Interpreter 2	No Support	CAI Support	No Support	CAI Support
Interpreter 3	No Support	CAI Support	No Support	CAI Support
Interpreter 4	No Support	CAI Support	No Support	CAI Support
Interpreter 5	No Support	CAI Support	No Support	CAI Support
Interpreter 6	No Support	CAI Support	No Support	CAI Support
Interpreter 7	No Support	CAI Support	No Support	CAI Support
Interpreter 8	No Support	CAI Support	No Support	CAI Support
Interpreter 9	No Support	CAI Support	No Support	CAI Support
Interpreter 10	No Support	CAI Support	No Support	CAI Support
Interpreter 11	No Support	CAI Support	No Support	CAI Support
Interpreter 12	No Support	CAI Support	No Support	CAI Support

3.4. PROCEDURE

The procedure involved the following procedure: Selecting, modifying and recording the four English texts (stimuli), dividing the texts into units of meaning, training participants on the Sight-Terp tool, conducting the repeated tests, assessing the quality of interpretation, and analysing the results through quantitative and qualitative methods. The descriptive outline of the procedure is as follows:

1. Four English texts were selected as the material for the study, which were then modified and shortened as described in section 3.3.1 to meet the objectives of the study. The content validity of the texts was ensured by using readability metrics (see Table 3.).

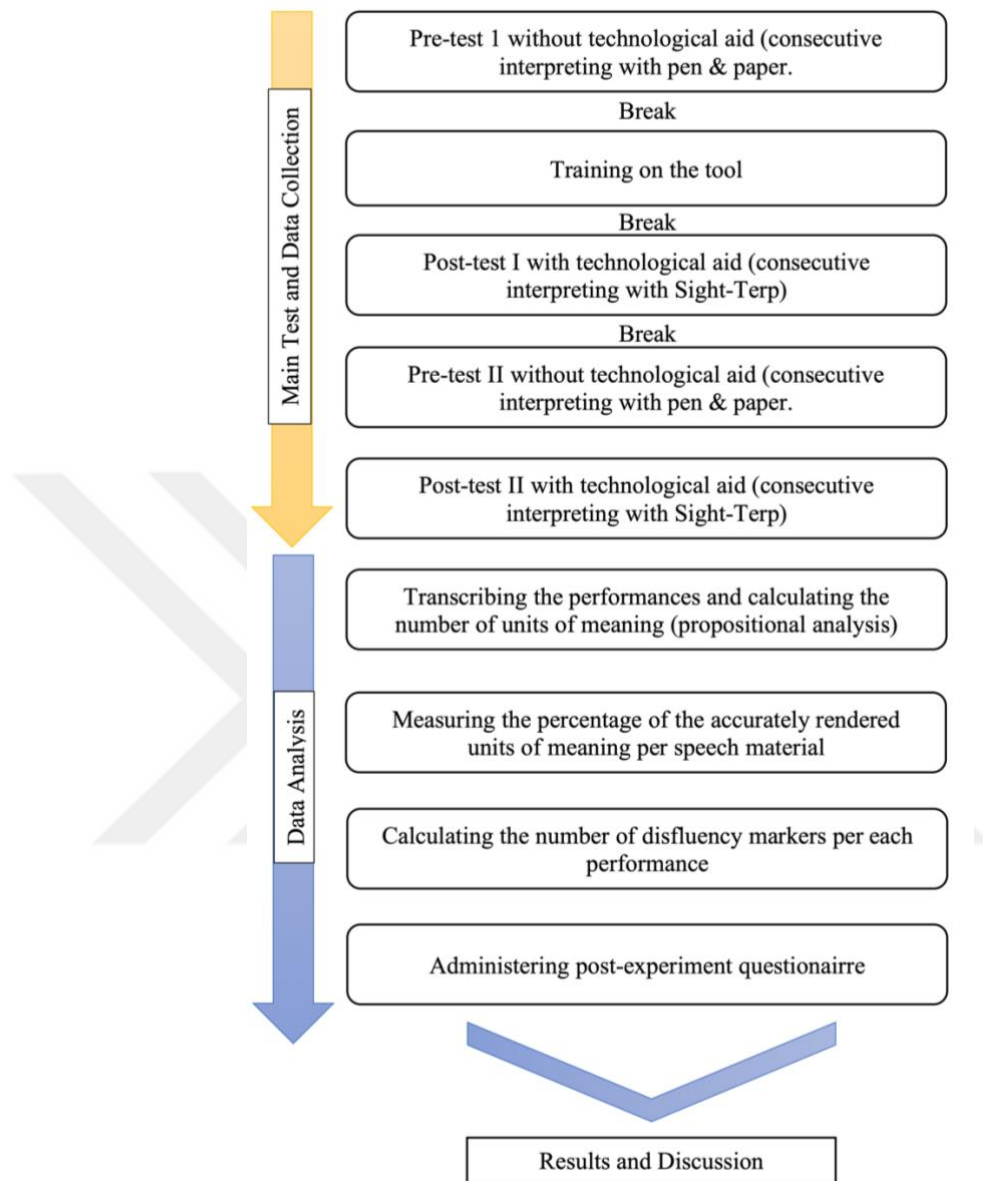
2. Speeches were recorded in a soundproof environment by a native speaker of English, at a reading speed that was natural and clear.
3. The recorded texts in both languages were divided into units of meaning, which were used as the basis for the consecutive interpreting tasks.
4. Participants were invited to the room where the experiment was conducted. The room is equipped with a table, chair, note-pad, pen, 11-inch Apple Ipad Pro (to run the tool Sight-Terp), and computer with speakers (to play the speeches). Upon arrival, participants were informed about the study's objectives, the procedure they would follow, and their rights as voluntary participants. They were then asked to sign a voluntary participation form to confirm their understanding and consent to participate in the study. With this formality completed, the experiment began.
5. A pre-test was conducted with the Speech A1 (Earthquakes in Japan). The participant invited for the experiment session was asked to interpret it into Turkish in consecutive mode with note-taking using pen and paper.
6. The participants were then provided with training on the use of the Sight-Terp. Features such as ASR, real-time speech translation, named entity recognition, and automatic segmentation of a speech are briefly introduced. Each participant is allowed to spend time using the tool and to speak into it with their own voice.
7. After 30 minutes, participants were then given a post-test, which involved interpreting Speech A2 into Turkish in consecutive mode using the Sight-Terp ASR-enhanced CAI tool.
8. The experiment is repeated with another pre-test and post-test using the other two materials. Speech B1(Violence Against Woman 1) without technological aid and Speech B2 (Violence Against Woman 2) with Sight-Terp.
9. The quality of the interpretation was assessed based on two criteria: accuracy and fluency. The accuracy was calculated as the percentage of the accurately rendered

"units of meaning" in each performance. For the two tests, wilcoxon Signed-Rank test was applied to understand the significant difference between two performances in two conditions, without technological aid and with Sight-Terp. Fluency on the other hand was measured by calculating the frequency of the disfluency markers such as overall frequency of disfluencies, false starts, frequency of filled pauses, filler words, whole-word repetitions, broken words, and incomplete phrases (Lickley, 2015).

- 10.** A follow-up qualitative survey was conducted to obtain comparative responses and perceptions on tool usage.
- 11.** Finally, the results were analyzed and discussed to evaluate the performance of the participants in consecutive interpreting tasks with and without the use of the ASR-enhanced CAI tool Sight-Terp, and to identify the benefits and challenges of incorporating ASR technology in the consecutive interpreting process.

Based on the procedure outlined above, the following chart delineates the steps taken within the scope of the study.

Chart 1. The procedure followed in the study



3.4.1. Training

The training phase constituted an integral part of the experiment, primarily aiming to familiarize the participants with the various functionalities of Sight-Terp. The process was designed to ensure that all participants were adequately equipped to interact and engage with the tool's features effectively.

The training session was structured to be comprehensive yet concise, which took about 35 minutes per participant. Initially, participants were introduced to Sight-Terp's critical features. This part of the session was dedicated to demonstrating how these features operate, explicating their practical implications in the context of the interpreting task, and offering guidance on the circumstances under which these features could be used.

Following the instructional component of the session, participants were encouraged to interact directly with Sight-Terp for a better understanding of the tool. This hands-on approach, including the participants trying out the tool using their own voice, skimming through the main interface, and physically engaging with the system, was crucial in terms of enabling participants to shift from passive learning to active application. This practical experience provided an opportunity for participants to explore and navigate through the intricacies of Sight-Terp since active engagement is during the testing phase.

The training phase aimed to minimize the learning curve associated with the use of Sight-Terp, ensuring that any performance outcomes observed during the testing phase could be attributed to the tool's impact rather than a participant's lack of familiarity with the tool.

3.4.2. Preliminary test

A pilot study was carried out at İstanbul Yeni Yüzyıl University, during the months of October and November 2022 with the aim of mapping the design and implementation of the main study and ensuring that it is well-suited to answer the research questions. The experimental design for the pilot study included a combination of process (partly) and product-oriented data collection methods. The primary objective of this study was to validate the stimuli designed for data collection. The approach taken was not without limitations, which were identified and modifications to the experimental design were made for the main study in preparation.

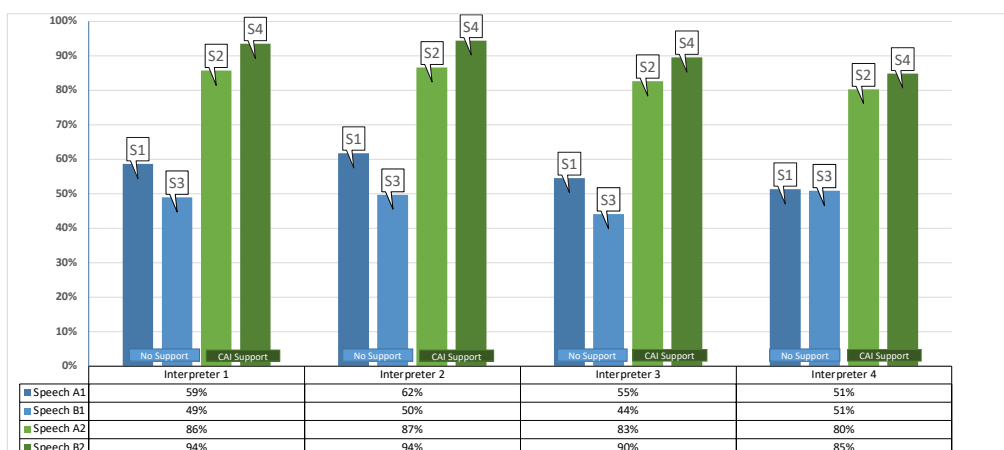
For the preliminary study, a small sample of four participants was used to evaluate the effectiveness of the test design in consecutive interpreting tasks. Of four participants, two are fresh graduates of translation and interpreting studies with a special focus on interpreting while the other two, similar to the participants in the main study, are senior

TIS students. All of them have successfully passed consecutive interpreting and note-taking courses during their undergraduate education. The age of the participants is between 24 and 21.

In the preliminary test, they were asked to interpret two speeches in traditional consecutive mode (using pen and paper) and two speeches using the Sight-Terp tool in sight-consecutive modality (using tablet). All participants were trained on how to use the tool prior to the experiment, using an 11-inch Apple iPad Pro. This pilot study also highlighted the potential benefits of incorporating the Sight-Terp tool in consecutive interpreting tasks, as evidenced by the increase in accurately rendered units of meaning when participants used the tool (Figure 14).

Three out of four participants reported that the ASR output does not fit the screen so they have to scroll constantly. Therefore, an “enlarge” button is added to the interface which expands the screen. The improvements made to the tool, experimental design and the positive outcomes from the preliminary study provided a solid foundation for the main study, which aims to further investigate the effectiveness of the Sight-Terp tool in consecutive interpreting tasks and answer the research questions more thoroughly. The preliminary study also provided valuable insights into the effectiveness of the stimuli materials and the overall experimental design, confirming that the chosen materials posed an adequate level of difficulty for the participants.

Figure 14. *The comparable results of the preliminary test: complete renditions of meaning units in %*



3.5. DATA ANALYSIS TECHNIQUES

While the administration of the test was underway, the participants' voices were recorded. The audio data gathered is subjected to a data analysis for the study, based on two variables: accuracy and fluency. For the analysis of the performances, accuracy was measured on the basis of a propositional analysis. In the context of this study, the four materials used are chunked into units of meaning (Seleskovitch, 1989) and the total number of each is calculated for further comparison. Units of meaning are defined by Seleskovitch as parts of meaning that appear at irregular intervals in the minds of those who consciously listen to understand speech (1989). The whole of semantic units is the synthesis of semantic information in the text. The units of meaning represent the structural meaning of a sentence and can be broken down into smaller elements. This method (involving propositional analysis techniques) mainly focuses on the semantic aspect of interpreting performance and is used by many researchers to assess the quality of interpreting (Dillinger, 1994; Tommola and Heleva, 1998; Orlando, 2014).

In the example given below, it can be mentioned that there are four units of meaning.

Dear participants,

In our talk today (1), we will talk (2) about the hygiene problems after the earthquake (3) and the steps that can be taken (4).

After the experiment, the total number of accurately rendered units of meaning was calculated for each participant to assess their accuracy rate. This was done by dividing the total number of accurately rendered units by the total number of units of meaning in the material and multiplying by 100 to get the percentage of correctly rendered units. The aggregated accuracy rates of the participants were then compared in the context of the pre-test and post-test to evaluate the effectiveness of the tool in enhancing interpreting accuracy.

A high level of granularity was applied in identifying each linguistic unit as a meaning unit, with subsequent analysis focusing on whether similar or complete transfer of these

units occurred in the target text. It should be noted that in line with some interpreting strategies applied during translation, semantic units may be subject to subtraction, addition, substitution and errors, consciously or unconsciously. While determining accurate transfer of the units of meaning in the target text, clauses or phrases with additions and deletions that would change the meaning expressed in the source text were not counted as units of meaning. Additionally, re-wordings and additions that did not negatively affect the context of the text and the intention of the speaker were not taken into account in denomination of the unit as a meaning unit.

Fluency is seen by some scholars as a measure of speech smoothness and continuity, while others perceive it as the interplay of temporal speech variables, such as pause length and uninterrupted speech runs, with factors like “voice clarity, enunciation, and speaker confidence” (Freed, 2000, p. 261). In interpreting studies general consensus on fluency is that speech rate, pauses, hesitations, lengthened syllables, repetitions, self-corrections, and false starts are units of the prosodic feature of a speech that affect fluency. In this study, in order to measure fluency, disfluency markers were analysed, including the overall frequency of disfluencies, false starts, filled pauses, filler words, whole-word repetitions, broken words, and incomplete phrases (Lickley, 2015). The number of occurrences of these markers in the participants' performance was calculated and aggregated. The total number of disfluencies was counted for each participant in both the pre-tests and post-tests. Then, the frequencies are compared in the contexts of two tests involving two conditions (with and without Sight-Terp) are outlined in the graph.

CHAPTER 4

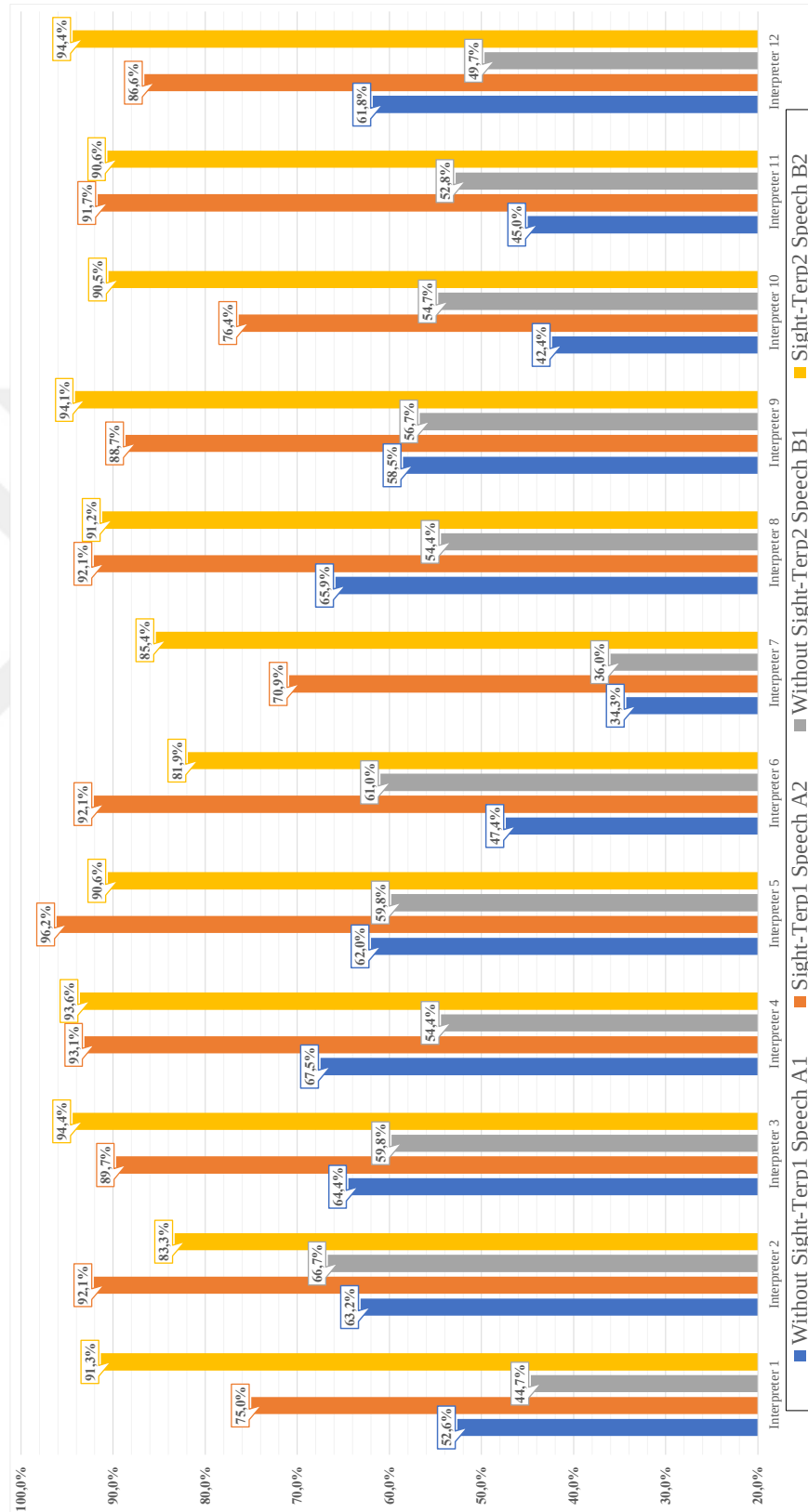
FINDINGS AND DISCUSSION

In this chapter, the findings of our study are presented, which aimed to investigate the impact of using the Sight-Terp tool, a web-based ASR-enhanced CAI tool, on interpreting accuracy and fluency among trainee interpreters ($n = 12$). To assess the effect of the technological aid on interpreting accuracy, run charts and the Wilcoxon Signed-Rank Test allowed us to examine the within-subject effect of the condition (No Technological Aid vs. Sight-Terp Aid) across four different measurements (No Technological Aid in SpeechA1, Sight-Terp in Speech A2, No Technological Aid in SpeechB1, and Sight-Terp in SpeechB2). Additionally, the number of occurrences of disfluencies in the performances is analysed manually. At the end of the experiment, participants filled in a post-experiment questionnaire where the interaction of users with the Sight-Terp tool and their general views, experiences, and suggestions are collected. Through the analysis of the data collected from our 12 participants, it is aimed to provide a comprehensive understanding of the extent to which the use of Sight-Terp can enhance interpreting accuracy and performance in CI.

4.1. FINDINGS AND DISCUSSION RELATED TO THE ACCURACY DIFFERENCES

This section presents the results and analysis of the experiment conducted to investigate the impact of the Sight-Terp tool on interpreting accuracy among interpreters. As described in the methodology chapter, the experiment followed a repeated measures design, and participants were asked to interpret two sets of speeches (SpeechA and SpeechB) both with and without the Sight-Terp tool (Speech A2 and B2 with the tool and Speech A1 and B1 without). Figure 15 is the run chart, which shows the percentage of the accurately rendered units of meaning in four of the speeches across all participants.

Figure 15. The comparable results of the main test: complete renditions of units of meaning in %.



Based on the data shown in the Figure 15, non-parametric Wilcoxon Signed-Rank Test was employed. The Wilcoxon Signed-Rank Test yielded a W value of 78.00, with a corresponding significance level of $p = 0.002$ ($n = 12$). Given that the significance level is below the conventional threshold of 0.05, it can be inferred that there is a statistically significant difference between the interpretation accuracy when the Sight-Terp tool was utilized and when no technological aid was used.

Further, conditional analysis across all values show that the interpretation accuracy was significantly higher in all Sight-Terp condition compared to the no-aid condition, with mean values of 87.05 (with Sight-Terp test 1) and 90.10 (With Sight-Terp test 2) and 55.41 (without Sight-Terp test 1) and 54.22 (without Sight-Terp aid test 2).

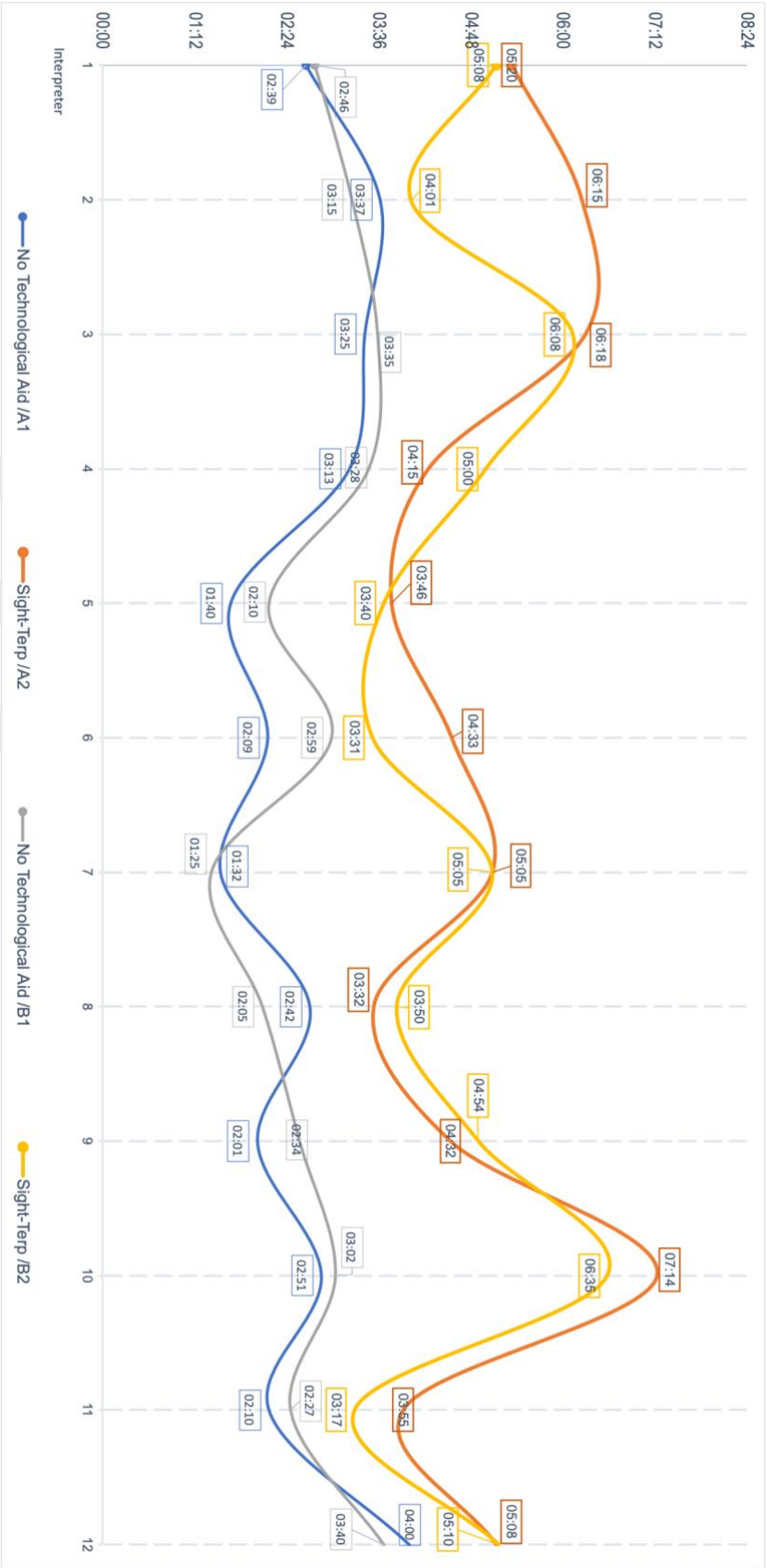
The effect size, as measured by $r = 0.882$, was found to be indicative of a large, further substantiating the significant influence of the Sight-Terp tool on the accuracy of interpretation.

To sum up, all twelve interpreters generally have higher accuracy in the "Sight-Terp" condition (when they used Sight-Terp) compared to the "without Sight-Terp" condition. Post-hoc analyses using pairwise comparisons can be conducted to further explore the differences between the conditions.

4.2. FINDINGS AND DISCUSSION RELATED TO THE FLUENCY DIFFERENCES

Figure 16 shows a run chart with the durations of each performance with and without Sight-Terp. A cursory examination of the data reveals that participants, on average, spent more time interpreting when they were provided with the Sight-Terp tool and comparatively took less time to deliver their interpretations with pen and paper.

Figure 16. The durations of the performances (in minutes and seconds)



It is observed that the higher accuracy in performances with Sight-Terp (as mentioned above) was accompanied by a longer interpretation time. Furthermore, this trade-off brought about a higher occurrence of disfluency markers.³⁶ As shown in Table 7, participants tend to show more disfluencies in the performances they used Sight-Terp.

Table 7. Instances of Disfluency Markers per Participant

Participant	No Tech. Aid	Sight-Terp	No Tech. Aid	Sight-Terp
	Speech A1	Speech A2	Speech B1	Speech B2
Interpreter 1	↓ 10	↑ 20	↓ 11	↘ 12
Interpreter 2	↓ 11	↑ 29	↓ 14	↓ 13
Interpreter 3	↓ 16	↑ 25	↓ 17	↘ 19
Interpreter 4	↓ 16	↘ 19	↘ 22	↑ 27
Interpreter 5	↓ 7	↑ 8	↓ 7	↑ 8
Interpreter 6	↗ 13	↑ 15	↓ 10	↘ 12
Interpreter 7	↓ 13	↗ 22	↘ 16	↑ 28
Interpreter 8	↗ 9	↑ 10	↓ 8	↘ 9
Interpreter 9	↓ 5	↘ 9	↘ 11	↑ 17
Interpreter 10	↓ 12	↑ 29	↘ 19	↑ 38
Interpreter 11	↗ 9	↑ 10	↘ 8	↓ 7
Interpreter 12	↓ 11	↑ 18	↘ 13	↘ 14

For the first test (With Sight-Terp Speech A1 vs No Tech. Aid Speech A2), the Z score is -3.065. The associated p-value (Asymp. Sig. (2-tailed)) is .002. For the second test (With Sight-Terp Speech B1 vs No Tech. Aid Speech B2), the Z score is -2.546. The associated p-value (Asymp. Sig. (2-tailed)) is .011, which is also less than the conventional alpha level of 0.05. Both tests show a statistically significant difference between the population mean ranks of Sight-Terp usage condition and pen&paper condition in the context of disfluency markers.

³⁶ As specified in the methodology part of this study, this comparison is conducted on an intra-individual basis, specifically by comparing each participant's performance with and without the use of technological assistance.

The higher disfluency rate observed in the Sight-Terp condition could be due to the increased cognitive demand, as interpreters not only had to process the spoken input but also had to manage the written text provided by the tool. Moreover, the presence of dual references (MT+ASR), to which the interpreters could resort in instances of minor inaccuracies or hesitation, may have also exerted a significant impact. This process could potentially interrupt the flow of interpretation, leading to more instances of false starts, filled pauses, filler words, whole-word repetitions, broken words, and incomplete phrases. In respect of longer durations, it is important to note that the higher rates of disfluency in Sight-Terp usage might not be the sole reason for longer durations of performances. In fact, having access to the whole ASR-generated source text coupled with MT might have compelled the interpreters to be more meticulous and complete, hence spending more time to deliver comprehensive interpretations. Therefore, it is evident that a high number of occurrences of disfluency accompanied by full content availability made the interpretations (with Sight-Terp) last even longer.

4.3. Post-experiment Questionnaire Results

The questionnaire aims to gain insights into participants' experiences and opinions about the Sight-Terp. The survey consists of eight questions, covering aspects such as overall experience, ease of use, perceived impact on performance, the reliability of the automatic speech recognition (ASR) function, specific output options used, and willingness to use the tool in future professional contexts. The survey employs a mixture of Likert scale, multiple-choice, and open-ended questions to capture a comprehensive understanding of participants' experiences and opinions.

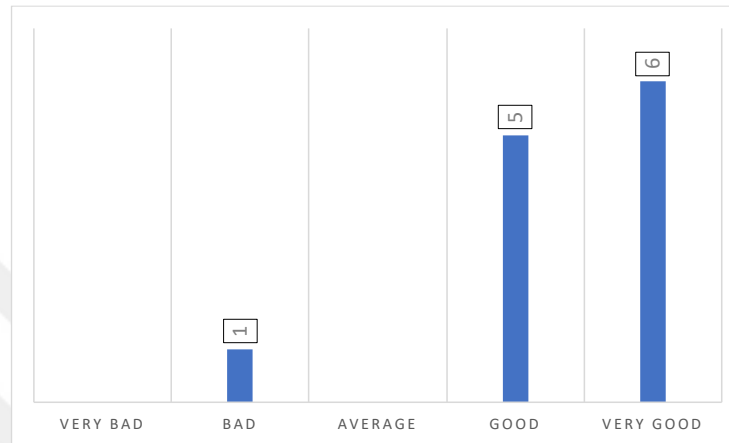
1. How would you evaluate your experience with the Sight-Terp tool?

Very Negative 1--5 Very Positive

Participants' overall experiences with Sight-Terp are mixed, as indicated by their responses to the open-ended question, "How would you evaluate your experience with Sight-Terp?". The participants were asked to evaluate their experience with a likert scale: Very Negative 1--5 Very Positive (Very Bad-Very Good). Detailed analysis of these

responses can provide further insight into the specific strengths and weaknesses of the tool from the users' perspectives. Figure 17 shows that the majority of the participants have positive experiences using Sight-Terp.

Figure 17. The answers to the question “How would you evaluate your experience with the Sight-Terp tool?”

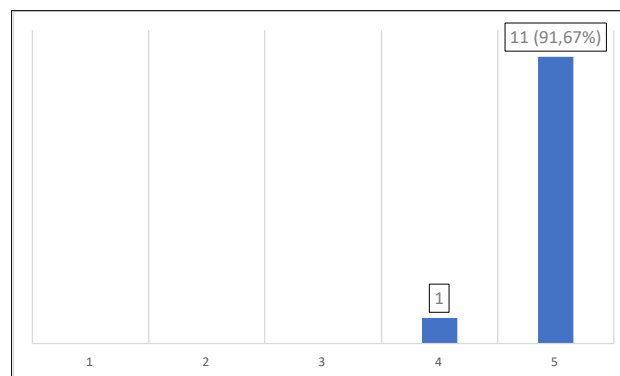


2. “I think the Sight-Terp tool is easy to use.”

Strongly disagree 1--5 Strongly agree

Regarding ease of use, participants responded to the Likert item, "I think that Sight-Terp is an easy-to-use tool." Responses to this question, depicted graphically in Figure 18, indicate the participants' perceived ease or difficulty in using Sight-Terp. Respondents' (11/12) general view is that Sight-Terp is easy to use.

Figure 18. The answers to the Likert item “I think the Sight-Terp tool is easy to use.”

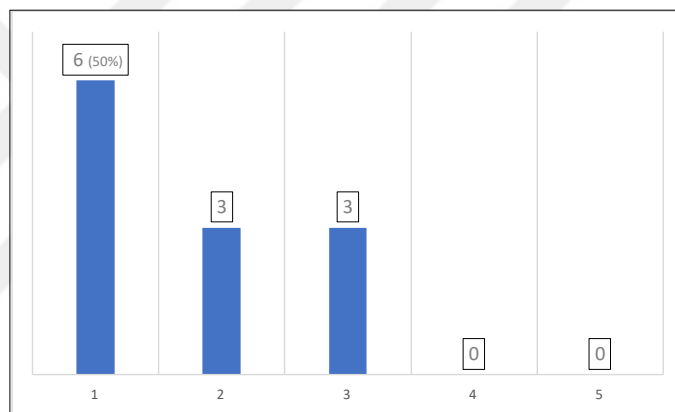


3. “Using automatic speech recognition during the consecutive interpreting task negatively affected my performance.”

Strongly disagree 1--5 Strongly agree

ASR on interpreting performance was assessed through the question "Did using ASR during consecutive interpreting tasks negatively impact your performance?" The distribution of responses, shown in Figure 19, reveals the participants' views on how ASR affected their interpreting performance.

Figure 19. The answers to the Likert item “Using automatic speech recognition during the consecutive interpreting task negatively affected my performance.”



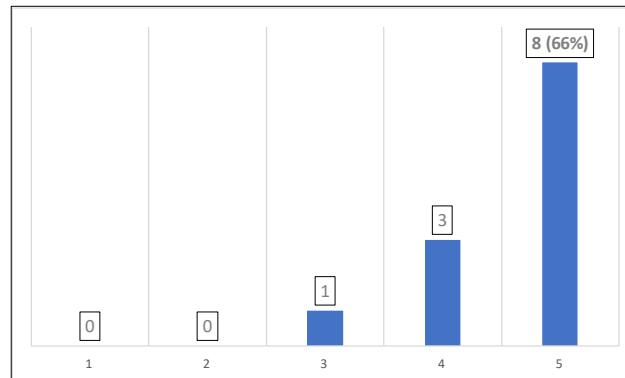
While most of the participants report no negativity of ASR on their performances, 3 out of 12 participants reported uncertainty on ASR having any effect on their performances partly or fully.

4. “I think the features in Sight-Terp contributed to my consecutive interpreting performance.”

Strongly disagree 1--5 Strongly agree

Similar to question 3, Figure 20 illustrates the distribution of responses, shedding light on how the participants perceived the tool's impact on their performance.

Figure 20. The answers to the Likert item “I think the features in Sight-Terp contributed to my consecutive interpreting performance.”



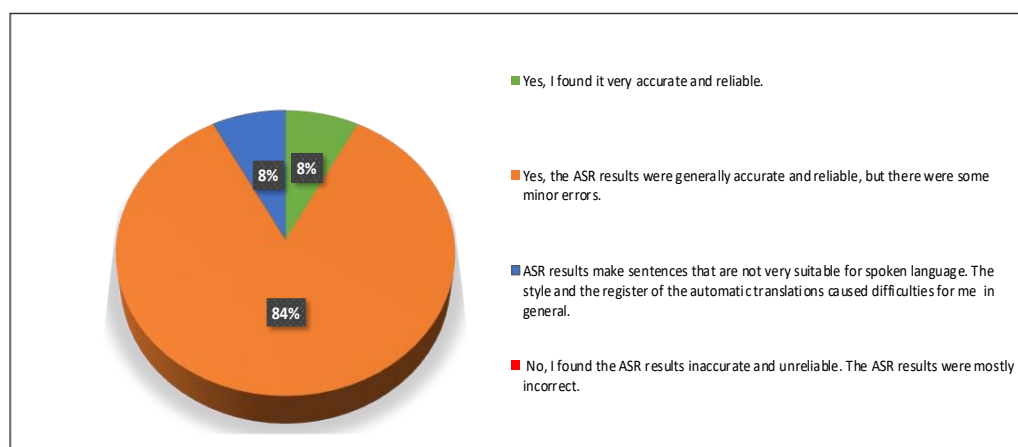
11 participants reported an observable contribution of features in Sight-Terp over their performances, while one participant stays neutral. In the last part of the questionnaire where participants are asked to provide their general opinions and recommendations, the participant giving “3” says that she/he benefited greatly from the reference texts and completed some sentences through on sight translation, some by reading them completely from the (machine) translation, and some by looking at both and combining them in an appropriate sentence. However, she/he thinks that she/he couldn't show the desired performance due to not being able to suppress the urge to use everything she/he saw on the screen.

5. Do you think the automatic speech recognition (ASR) function in Sight-Terp is accurate and reliable? (Multiple-choice)

Options:

- Yes, I found it very accurate and reliable.
- Yes, the ASR results were generally accurate and reliable, but there were some minor errors.
- I have no opinion.
- No, I found the ASR results inaccurate and unreliable. The ASR results were mostly incorrect.
- Other

Figure 21. The answers to the question “Do you think the automatic speech recognition function in Sight-Terp is accurate and reliable?”



When asked, "Do you think that the automatic speech recognition (ASR) function in Sight-Terp is accurate and reliable?" all participants provided one response except one who preferred to give his/her own response. These responses, depicted in Figure 21, indicate the participants' perceptions of the accuracy and reliability of the ASR function in Sight-Terp are slightly positive. However, the general view is that ASR results (as well as MT) are not fully reliable since only one participant reported full confidence. Although word-error-rate of the materials is satisfyingly good, participants' general view of the accuracy as "not so reliable" might stem from the fact that any occurrence of error might have made participants act with suspicion towards the full output.

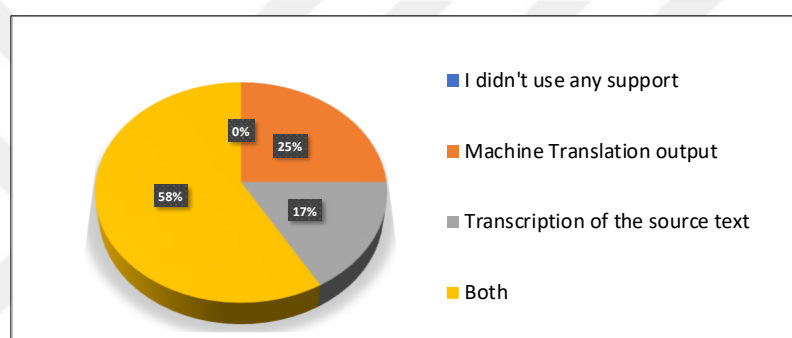
6. Which automatically generated output did you use for support during consecutive interpreting? (Multiple-choice)

Options:

- I didn't use any support.
- Machine translation output
- Source text output
- Both

In response to the question sixth question, "Which automatically generated output did you use for support during consecutive interpreting?" participants gave various responses. These responses, represented in Figure 22, provide insight into which features of Sight-Terp were most used and found most useful by the participants. Since the cognitive process lying the task seems complex and the source text available for interpreters is based on two options (transcription or MT), the responses might be illuminating for their preference of use and the impact of their preference on the accuracy/fluency results.

Figure 22. The answers to the question “Which automatically generated output did you use for support during consecutive interpreting?”



Participants generally used both outputs (7/12). Out of three participants having selected the answer “Machine Translation output” here, only one reported that she/he thinks ASR results were very accurate and reliable in question 5. The other two have generally used the MT output though they did not rely on the results fully. When asked about the reason behind their choice of reference, these three participants reported that once they are more or less sure the translation is correct, they feel more fluent interpreting using the MT (through monolingual editing). Therefore, interpreting from the MT reference made them feel more fluent.

On the other hand, there are two participants (2/12) who did not refer to the MT but the source transcription while interpreting though they responded to question 5 as “Yes, the ASR results were generally accurate and reliable but there were some minor errors.”. When asked about the reason, participants using the source transcription output for reformulating their target text gave various answers implying that they did not ‘feel

actually interpreting' and felt uncomfortable if they try to read the automatic translation out loud and word by word. However, MT reference was discarded at all, since they reported that they occasionally looked up some words in the MT reference. Participants, having used both of the reference texts, indicate that they tend to use machine translation for long and complex sentences. In case they are unsure about any word or suspicious about the structure of the translated text, they can take the liberty to the source text immediately.

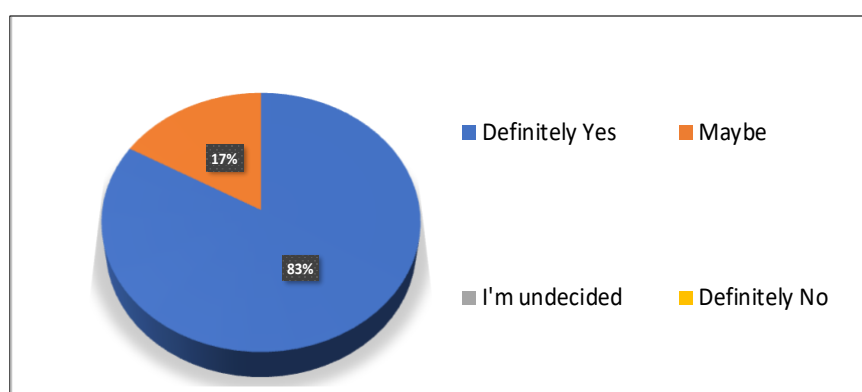
7. Would you use the Sight-Terp tool in your future professional life? (Multiple-choice)

Options:

- Definitely yes
- Maybe
- I'm undecided
- Definitely no

When asked, "Would you use the Sight-Terp tool in your future professional life?" participants' responses were varied. These responses, depicted in Figure 23, reveal how participants envisage the potential role of Sight-Terp in their future work.

Figure 23. *Answers to the question "Would you use the Sight-Terp tool in your future professional life?"*



Participants' general view on their future use is a big yes. Participants selecting "Maybe" (2/12) were particularly asked about the reason behind their decision. The common answer to this is that they want to see or use the tool on different types of assignments and settings to be completely sure. 4 out of 10 participants saying "Definitely Yes" reported that they would use Sight-Terp but they want to use it while still taking notes on a paper. As described in 2.4.1.4, Sight-Terp actually includes an option to take notes on a digital notepad using a stylus like Apple Pen or Samsung S Pen. However, due to the scope of this study, note taking simultaneously on the digital notepad was not allowed.

8. Is there a feature you would like to see in Sight-Terp? If yes, what is it? (Open-ended)

In the last part of the questionnaire, the participants were asked to provide feedback and whether there is any feature they would like to see in the tool. Several key themes and potential areas for improvement have emerged for Sight-Terp:

- **Segmentation Consistency:** Sometimes texts displayed on the interface are not segmented properly which makes MT get away from coherence in the context of the speech. These segmentation issues make the interpreter mentally edit the fragmented text and deliver it accordingly during interpretation. Respondents reported a few inconsistencies in the segmentation of the text, particularly when dealing with closely related elements and concepts such as "first of all," "secondly," etc. The problem generally occurs because of the ASR, which sometimes fails to predict the end of the sentence when the speaker gives a long pause in the sentence.
- **Segment Manipulation:** A manual segment merge or split feature was suggested, which could potentially contribute to better translation quality.
- **Automatic Scrolling:** Participants found having to manually scroll to see new tabs added during the speech to be inconvenient. Automatic scrolling as the speech unfolds could improve the user experience.

- **Ability to Post-edit:** Users suggested the addition of an editor for real-time quick corrections of minor errors in the machine translation. This post-editing can also be automatized by correcting the typos or minor errors in the ASR results through a language model or there might be suggestions below each problematic word.
- **Automatic Highlighting in Target Text:** Users expressed a desire for the ability to click on highlighted words or phrases to see their equivalents in the target language. In the current version of Sight-Terp, only source text has named entities highlighted. Some users suggested that countries, numbers, and percentages (named entities in general) should also be highlighted in the target text for easier reference. They also suggested that manually highlighting specific data such as years, proper names, etc., in both the source and target texts could improve the tool's utility.
- **Segment Lines:** Two participants suggested thin lines between each segment to easily distinguish between the segments while reading and not to confuse anything while navigating between two references.

The feedback from participants indicates areas where Sight-Terp is performing well, as well as areas for potential improvement. These findings can inform the future development of the tool, with the aim of enhancing its utility and usability for interpreters.

CONCLUSION AND RECOMMENDATIONS

In conclusion, this thesis has aimed to investigate the effectiveness and potential of the CAI tool Sight-Terp in enhancing interpreting performance in consecutive interpreting. The introductory chapter provided a comprehensive overview of the research objectives, significance, research questions, limitations, assumptions, and research definitions. Chapter two delved into the background and literature review, shedding light on the historical and etymological aspects of interpreting. The definition of CAI tools was provided, accompanied by examples of ASR-enhanced CAI tools available on the market. Furthermore, speech technologies and their integration into interpreting were examined, incorporating qualitative and quantitative data from various studies. The usage of technological solutions for consecutive interpreting was also highlighted, concluding with a detailed description of the proposed CAI tool, Sight-Terp. Chapter three elucidated the methodology employed in this study. The subsequent chapter, chapter four, presented the findings and discussions derived from the study. It analysed the accuracy and fluency differences in interpreting performance, considering the utilization of CAI tools. Furthermore, comprehensive feedback from users of the CAI tool was examined, providing insights into the user experience and perceptions of its effectiveness. From this point forward, the research questions raised at the beginning of this thesis will be addressed and recommendations for future research will be given.

Conclusion Regarding Research Questions

Q1. Does the use of the CAI tool Sight-Terp in consecutive interpreting lead to a significant improvement in the interpreting accuracy and performance of interpreters in a sight-consecutive modality compared to their performance without technological aid?

Through quantitative and small-scale qualitative analysis, it has been observed that the use of ASR-enhanced CAI tool Sight-Terp leads to a noteworthy improvement in the content accuracy of interpreting performances. The experiment showed that, in this sight-consecutive modality, participants exhibited increased precision in their renditions and a deeper engagement with the text when employing Sight-Terp, as compared to interpreting without technological aid (only pen and paper). Its use is indeed a substantial step forward

in promoting and ensuring quality in the integration of ASR into interpreting practice. However, as stated in the answer to the research question two below, it must be acknowledged that there was an increase in instances of disfluencies and an extension of interpretation duration when Sight-Terp was used. Despite these drawbacks, these factors did not significantly overshadow the observable enhancements in interpretation accuracy. Further research and development in Sight-Terp can possibly alleviate the aforementioned concerns and make it an even more effective aid in the interpreting process.

Q2. *Are there significant differences in the number of disfluencies (pauses, hesitations, repetitions, stuttering, false starts instances) between pre-test performances without CAI support and post-test performances with Sight-Terp support?*

The analysis of the data reveals that the instances of disfluencies, including pauses, hesitations, repetitions, stuttering, and false starts, were noticeably higher in post-test performances when participants used the Sight-Terp tool. This increase in disfluency markers suggests that the use of Sight-Terp may have influenced the flow of interpretation, potentially due to the cognitive load associated with processing additional information provided by the tool. In further studies, the extent and dimension of the cognitive load can be unveiled with robust empirical methods, since the number of occurrences of disfluency markers provide limited understanding on the underlying aspects.

As for the durations, participants consistently took longer to complete their interpretations when using the Sight-Terp tool compared to the no-aid condition. This trend suggests that the use of Sight-Terp may have an impact on the time required to deliver an interpretation, potentially due to a couple of reasons. As a result, while the Sight-Terp tool seems to enhance accuracy, it also appears to extend the time needed for the interpretation process. The extended time could be attributable to a variety of factors such as the additional/redundant information provided by the tool, unfamiliarity with the tool, and the need for interpreters (participants) to integrate this information into their work. For instance, interpreters may require additional time to read and process the text provided by Sight-Terp. Alternatively, the increased accuracy provided by Sight-Terp may have

encouraged interpreters to be more meticulous in their interpretation, thus increasing the time taken. The extended time and increased disfluencies could also be attributed to the participants familiarizing themselves with the new tool and integrating it into their workflow.

Q3. *How do users interact with the tool Sight-Terp? Do its interface design and ergonomic features meet the required standards for efficient and effective interpretation?*

Based on the comprehensive feedback from the study participants, it is evident that Sight-Terp has significant potential to be an effective tool in aiding interpreting. The majority of the users found the platform easy to use and using the tool on a tablet prevented unfamiliarity, suggesting that its interface design and ergonomic features, to some extent, meet the required standards for efficient and effective interpretation. However, some challenges were reported, particularly regarding minor errors in ASR results and inconsistencies in the segmentation of the chunks. Although most users reported a positive experience, there was some uncertainty about the impact of ASR on interpretation performance. This suggests that while the ASR is a valuable feature, its implementation could be optimized to reduce any perceived negative impact on performance. Additionally, participants feel that more experience with Sight-Terp would entail more familiarity. As such, empirical research with larger sampling would generate more user-centric data to diminish limitations.

The questionnaire also revealed that users found the functions available in Sight-Terp beneficial for their interpreting performance, highlighting the usefulness of the tool in supporting interpreters. However, the reliability and accuracy of ASR and MT results were viewed with some scepticism. This scepticism, albeit minor, underscores the need for further improvements in these features to increase user trust and confidence. Interestingly, the study found that users employed different strategies when utilizing Sight-Terp's automatically generated outputs for support during consecutive interpreting. This indicates that Sight-Terp allows for a certain level of flexibility and adaptability, allowing interpreters to work in a way that suits their individual preferences and strategies. More comprehensive studies investigating differences in the performances

among three different reference choices (only MT, only ASR-generated source transcript or both) can yield promising results about the disadvantages of each. Though almost half of the participants reported they sought support in both of the reference texts, it is unknown which one supported them most.

In terms of future usage, there is a strong inclination among participants to use Sight-Terp in their professional lives, suggesting its potential for widespread adoption. However, some users expressed a desire to test the tool in different assignments and settings before fully committing to its use, emphasizing the need for further validation of Sight-Terp in various professional contexts.

In conclusion, users interact with Sight-Terp in diverse ways, and while the tool is generally regarded as easy to use and beneficial for interpreting performance (on the side of accuracy as the graph on accuracy shows), there are areas, particularly in relation to ASR and MT results, where improvements could be made to enhance user experience and confidence.

Recommendations For Future Research

There are many opportunities for exploration and discovery in the nascent nature of CAI tool research. As more and more researchers turn their attention to this emerging field, it is likely that new insights will be uncovered. This will lead to the development of more advanced and sophisticated CAI tools. The potential impact of technology on the interpreting industry is significant, so it is imperative that scholars and researchers continue to explore this field with a high level of curiosity and rigour. The lack of a clear definition and research agenda presents challenges for a thorough mixed-methods methodology. However, the emerging nature of the field is undeniable and its potential impact on the interpreting process is significant.

Considering a vast and detailed research area, there is a need for a more empirical study of CAI tool research, from need analysis to tool performance, from its impact on users to reflections on the interpreting community. In light of this study, the following recommendations might be listed:

1. Future research can be conducted on a larger scale with a more diverse participant pool but with different language pairs other than Turkish-English as language
2. Another important avenue for future research can be a similar study which scrutinizes the cognitive processes involved in using the CAI tool and its impact on cognitive load. Triangulating eye-tracking data and transcription analysis could provide insight into how users interact with the tool and identify areas for improvement.
3. Further studies can investigate the impact of Sight-Terp and other ASR-enhanced CAI tools on interpreter training and education and explore potential ways to incorporate these technologies into interpreting curricula in different graduate levels and countries.
4. The digital note-pad feature of Sight-Terp has not been used within the scope of this study. Future studies might unveil the interoperability of digital note-pad and ASR together using a tablet.
5. As this study is conducted with relatively novice interpreters (students) on mock-up settings (pre-recorded speeches), a study with larger sample (including professional interpreters) on real-life scenarios may unveil other potentials of ASR and CAI in consecutive interpreting.
6. The efficacy level of the provision of two references and are yet to be tested. By analyzing interpretations at the textual level, it would be possible to discern the instances where participants deviate from the source transcription and rely on machine translation. This small-scale corpus work could reveal areas of difficulty as well as the coping strategies employed by the interpreters.

BIBLIOGRAPHY

- Abbasbeyli, E. (2015, May 22). The Role of Dragomans in the Ottoman Empire. AIIC Blog. <https://aiic.net/p/7219>.
- Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, 6(1), 43-67.
- Albarino, S. (2023, January 4). Research on speech-to-speech translation is booming. Slator. <https://slator.com/research-on-speech-to-speech-translation-booming/>
- Altieri, M. (2020). Tablet interpreting: Étude expérimentale de l'interprétation consécutive sur tablette. *The Interpreters' Newsletter*, 25, 19-35. <http://doi.org/10.13137/2421-714X/31235>
- Andres, D. (2002). *Konsekutivdolmetschen und Notation*. FASK.
- Andres, D. (2012). History of interpreting. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2512–2521). Blackwell.
- Andres, D. (2015). Consecutive interpreting. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 84–87). Routledge.
- Ann T., & John T. (1983). *The Nuremberg Trial*. Macmillan.
- Arumí, J.L. & Sánchez-Gijón, P. (2019). La toma de notas con ordenadores convertibles en la enseñanza-aprendizaje de la interpretación consecutiva. Resultados de un estudio piloto en una formación de master. *Revista Tradumàtica Technologies de la Traducció*, 17, 128-152. <https://doi.org/10.5565/rev/tradumatica.234>
- Baigorri-Jalón, J. (2014). *From Paris to Nuremberg: The birth of conference interpreting* (H. Mikkelsen & B. S. Olsen, Trans.). John Benjamins.
- Baigorri-Jalón, J. (2015). The history of the interpreting profession. In H. Mikkelsen & R. Joudenais (Eds.), *The Routledge Handbook of Interpreting* (pp. 11–28). Routledge.
- Bérard, A., Pietquin, O., Servan, C., & Besacier, L. (2016). Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation. In *NIPS Workshop on end-to-end learning for speech and audio processing*, Barcelona, Spain.
- Biagini, G. (2015). *Glossario cartaceo e glossario elettronico durante l'interpretazione simultanea: uno studio comparativo (Master's thesis)*. Università di Trieste.
- Braun, S. (2019). Technology and interpreting. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 271-288). Routledge.
- Čeňková, I. (2015). Sight interpreting/translation. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 374–375). Routledge.

- Cheung, A., & Tianyun, L. (2018). Automatic speech recognition in simultaneous interpreting: A new approach to computer-aided interpreting. In *Proceedings of Ewha Research Institute for Translation Studies International Conference*. Ewha Womans University.
- Ciobanu, D. (2014). Of Dragons and Speech Recognition Wizards and Apprentices. *Revista Tradumàtica*, (12), 524–538.
- Ciobanu, D. (2016). Automatic Speech Recognition in the Professional Translation Process. *Translation Spaces*, 5(1), 124–144.
- Ciobanu, D., & Secară, A.A. (2019). Speech recognition and synthesis technologies in the translation workflow. In J. W. Schwieter & A. Ferreira (Eds.), *The Routledge Handbook of Translation and Technology* (pp. 91-103). Routledge.
- Ciobanu, D., Ragni, V., & Secară, A. (2019). Speech Synthesis in the Translation Revision Process: Evidence from Error Analysis, Questionnaire, and Eye Tracking. *Informatics*, 6(4), 51. MDPI AG. <http://dx.doi.org/10.3390/informatics6040051>.
- Corpas Pastor, G. (2021). Interpreting and Technology: Is the Sky Really the Limit? In *Proceedings of the Translation and Interpreting Technology Online Conference TRITON 2021*.
- Cui, L., Wu, Y., Liu, J., Yang, S., & Zhang, Y. (2021). Template-based named entity recognition using BART. arXiv preprint arXiv:2106.01760.
- Davis, R., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digit. *Journal of the Acoustical Society of America*, 24, 637.
- Defrancq, B., & Fantinuoli, C. (2021). 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), 73-101.
- Delisle, J., & Woodsworth, J. (2012). *Translators through history* (Revised Edition). John Benjamins Pub. Co.
- Deng, L., & Li, X. (2013). Machine learning paradigms for speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5), 1060-1074.
- Desmet, B., Vandierendonck, M., & Defrancq, B. (2018). Simultaneous interpretation of numbers and the impact of technological support. In *Interpreting and Technology* (pp. 13–28). Zenodo.
- Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't?. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (pp. 155-189). John Benjamins.

- Diriker E. (2018). Sözlü Çeviri: Tarih, Eğitim ve Araştırmalar. In Diriker E. (Eds.) *Türkiye’de Sözlü Çeviri: Eğitim, Uygulama ve Araştırmalar*, İstanbul, Scala (pp. 13).
- Diriker, E. (2005). *Konferans çevirmenliği: Güncel Uygulamalar Ve Araştırmalar*. Scala yayıncılık.
- Doğan, A. (2022). *Sözlü Çeviri Çalışmaları ve Uygulamaları*, Siyasal Publishing.
- Dragsted, B., & Gorm Hansen, I. (2007). Speaking your translation: Exploiting synergies between translation and interpreting. In F. Pöchhacker, A. L. Jakobsen, & I. M. Mees (Eds.), *Interpreting Studies and Beyond: A Tribute to Miriam Shlesinger* (pp. 251–274). Copenhagen: Samfundslitteratur Press.
- Dragsted, B., Mees, I. M., & Hansen, I. G. (2011). Speaking your translation: students’ first encounter with speech recognition technology. *The International Journal for Translation & Interpreting Research*, 3(1), 10–43.
- Drechsel, A., & Goldsmith, J. (2016). Tablet Interpreting: the evolution and uses of mobile devices in interpreting. *In Proceedings of CUITI Forum 2016*.
- Duquenne, P., Gong, H., Dong, N., Du, J., Lee, A., Goswami, V., Wang, C., Pino, J. M., Sagot, B., & Schwenk, H. (2022). SpeechMatrix: A large-scale mined corpus of multilingual speech-to-speech translations. *arXiv preprint*. arXiv:2211.04508.
- EABM. (2021b). [Survey]. Ergonomics for the artificial boothmate. <https://www.eabm.ugent.be/survey/>
- Ergonomics for the Artificial Booth Mate Project (EABM). (2021a). Cai. <https://www.eabm.ugent.be/cai/>.
- Erik F. Tjong Kim Sang & Fien De Meulder. (2003). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. *In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003* (pp. 142-147).
- Fantinuoli, C. (2009). InterpretBank: Ein Tool zum Wissensmanagement für Simultandolmetscher. In W. Baur, S. Kalina, F. Mayer, & J. Witzel (Eds.), *Übersetzen in die Zukunft: Herausforderungen der Globalisierung für Dolmetscher und Übersetzer: Tagungsband der Internationalen Fachkonferenz des Bundesverbandes der Dolmetscher und Übersetzer e.V* (pp. 411-417).
- Fantinuoli, C. (2012). *InterpretBank—Design and Implementation of a Terminology and Knowledge Management Software for Conference Interpreters*. (Phd Dissertation). The University of Mainz.
- Fantinuoli, C. (2012). *InterpretBank: Design and implementation of a terminology and knowledge management software for conference interpreters* (Unpublished Phd Thesis). Johannes Gutenberg-Universität Mainz.

- Fantinuoli, C. (2016). InterpretBank. Redefining computer-assisted interpreting tools. In *Proceedings of the Translating and the Computer 38* (pp. 42-52). Geneva: Editions Tradulex.
- Fantinuoli, C. (2017a). Speech Recognition in the Interpreter Workstation. In *Proceedings of the Translating and the Computer 39* (pp. 25-34). Geneva: Editions Tradulex.
- Fantinuoli, C. (2017b). Computer-assisted preparation in conference interpreting. *The International Journal of Translation and Interpreting Research*, 9(2), 19-30.
- Fantinuoli, C. (2018a). Computer-assisted interpreting: challenges and future perspectives. In I. Durán & G. Corpas (Eds.), *Trends in e-tools and resources for translators and interpreters* (pp. 153-174). Leiden: Brill.
- Fantinuoli, C. (2018b). Interpreting and technology: The upcoming technological turn. In *Interpreting and Technology* (pp. 1-12). Zenodo.
- Fantinuoli, C. (2019). The technological turn in interpreting: The challenges that lie ahead. In W. Baur & F. Mayer (Eds.), *Übersetzen und Dolmetschen 4.0: Neue Wege im Digitalen Zeitalter* (pp. 334-354) BDÜ Fachverlag.
- Fantinuoli, C. (2022, September 22). Automatic Speech Recognition for interpreters: State-of-the-art and future prospective [Video]. UMassAmherst - Faculty and community seminar on interpreting studies and practice. YouTube. <https://youtu.be/c9uvGrP6A9A?t=1967>
- Fantinuoli, C., & Montecchio, M. (2022). Defining maximum acceptable latency of AI-enhanced CAI tools. *arXiv preprint arXiv:2201.02792*.
- Fantinuoli, C., & Pisani, E. (2021). Measuring the Impact of Automatic Speech Recognition on Number Rendition in Simultaneous Interpreting. In *Empirical Studies of Translation and Interpreting (1st ed.)*. Routledge.
- Fantinuoli, C., Marchesini, G., & Landan, D. (2022). Interpreter Assist: Fully-automated real-time support for Remote Interpretation. In *Proceedings of Translator and Computer 53 Conference*.
- Fantinuoli, Claudio & Bianca Prandi. 2021. Towards the evaluation of automatic simultaneous speech translation from a communicative perspective. In *Proceedings of the 18th International Conference on Spoken Language Translation* (pp. 245–254). Bangkok: Association for Computational Linguistics.
- Flerov, C. (2013, October 30). On comintern and hush-a-phone: Early history of simultaneous interpretation equipment. AIIC.net. https://aiic.org/document/893/AIICWebzine_2013_Issue63_5_FLEROV_On_Comintern_and_Hush-a-Phone_Early_history_of_simultaneous_interpretation_equipment_EN.pdf

- Fohr, D., Mella, O., & Illina, I. (2017). New paradigm in speech recognition: Deep neural networks. *Paper presented at the IEEE International Conference on Information Systems and Economic Intelligence*, Marrakech, Morocco.
- Freed B.F. (2000) Is fluency, like beauty, in the eyes (and ears) of the beholder. In H. Riggensbach (ed.), *Perspectives on Fluency*. (pp. 243-265). University of Michigan Press.
- Frittella, F. M. (2019). “70.6 billion world citizens”: Investigating the difficulty of interpreting numbers. *The International Journal of Translation and Interpreting Research*, 11(1), 79–99. <https://doi.org/10.12807/ti.111201.2019.a05>
- Frittella, F. M. (2023). *Usability research for interpreter-centred technology the case study of SmartTerp*. Language Science Press.
- Frittella, F. M. & Rodríguez, S. (2022). Putting SmartTerp to Test: A tool for the challenges of remote interpreting. *NContext*, 2(2), 137-166. <https://doi.org/10.54754/incontext.v2i2.21>
- Gaber, M., Corpas Pastor, G., & Omer, A. (2020). Speech-to-Text Technology as a Documentation Tool for Interpreters: a new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. *TRANS. Revista De Traductología*, 24, 263-281. <https://doi.org/10.24310/TRANS.2020.v0i24.7876>
- Gacek, M. (2015). *Softwarelösungen für DolmetscherInnen (Master's thesis)*. University of Vienna.
- Gaiba, F. (1998). *The Origins of Simultaneous Interpretation: The Nuremberg Trial*. University of Ottawa Press.
- Gile, D. (1985). Le modèle d’efforts et l’équilibre en interprétation simultanée [The effort model and balance in simultaneous interpreting]. *Meta*, 30(1), 44–48.
- Gile, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training*. John Benjamins.
- Gile, D. (1997). Conference interpreting as a cognitive management problem. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 163-176). Routledge.
- Gile, D. (1998). Conference and simultaneous interpreting. In M. Baker (Ed.), *Routledge Encyclopedia of Translation Studies* (pp. 40-45). Routledge.
- Gile, D. (1999). Testing the effort models’ tightrope hypothesis in simultaneous interpreting - A contribution. *Hermes - Journal of Linguistics*, 23, 153-172.
- Gile, D. (2001). The role of consecutive in interpreter training: A cognitive view. https://aiic.org/document/436/AIICWebzine_SepOct2001_1_GILE_The_role_of_consecutive_in_interpreter_training_A_cognitive_view_EN.pdf

- Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training*. John Benjamins.
- Gile, D. (2020). Forty years of effort models of interpreting: Looking back, looking ahead [Keynote lecture]. Japan Translation and Interpretation Forum 2020 organized by JACI – Japan Association of Conference Interpreters.
- Gile, D. (2023). The effort models and gravitational model clarifications and update. <http://doi.org/10.13140/RG.2.2.20178.43209>.
- Gillies, A. (2017). *Note-taking for consecutive interpreting: A short course*. Taylor & Francis Group.
- Goldsmith, J. (2017). A comparative user evaluation of tablets and tools for consecutive interpreters. In *Proceedings of Translation and the Computer 39* (pp. 39-50).
- Goldsmith, J. (2018). Tablet interpreting: Consecutive interpreting 2.0. *Translation and Interpreting Studies*, 13(3), 342-365. <https://doi.org/10.1075/tis.00020.gol>
- Goldsmith, J., & Holley, J. (2015). *Consecutive interpreting 2.0: The tablet interpreting experience (Unpublished MA Thesis)*. University of Geneva.
- Hamidi, M., & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 52(2), 276-289. <http://doi.org/10.7202/016070ar>
- Herbert, J. (1952). *Manuel de l'interprète: Comment on devient interprète de conférences*. Geneva: Librairie de l'Université.
- Hermann, A. (2002). Interpreting in antiquity. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 15-22). Routledge. (Original work published 1956)
- Hiebl, B. (2011). *Simultanes Konsekutivdolmetschen mit dem Livescribe Echo Smartpen. (Master's thesis)*. University of Vienna.
- Hitzel, F. (1995). *Enfants de langue et Dragomans: Dil Oğlanları ve Tercümanlar*. İstanbul: Yapı Kredi.
- Jones, R. (2002). *Conference Interpreting Explained*. St Jerome.
- Kade, O. (1968). *Zufall und Gesetzmäßigkeit in der Übersetzung [Chance and Regularity in Translation]*. Verlag Enzyklopädie.
- Lamberger-Felber, H. (2001). Text-oriented research on interpreting: Examples from a case study. *Hermes: Journal of Linguistics*, 26, 29–64.
- Lamberger-Felber, H. (2003). Performance variability among conference interpreters: Examples from a case study. In Á. Collados Aís, M. M. Fernández Sánchez & D.

- Gile (Eds.), *La evaluación de la calidad en interpretación: Investigación* (pp. 147–168). Granada.
- Lauterbach, E., & Pöchhacker, F. (2015). Interference. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 194–195). Routledge.
- Lederer, M. (1981). *La traduction simultanée – Expérience et théorie [Simultaneous interpreting - Experience and theory]*. Les Éditions Minard.
- Lickley, R. J. (2015). Fluency and Disfluency. In M. Redford (Ed.), *The Handbook of Speech Production* (pp. 445–474). Hoboken, NJ: John Wiley and Sons.
- Liu, A. H., Hsu, W., Auli, M., & Baevski, A. (2022). Towards end-to-end unsupervised speech recognition. *2022 IEEE Spoken Language Technology Workshop (SLT)*, 221–228.
- Liyanapathirana J. and Bouillon P. (2022) Integrating post-editing with Dragon speech recognizer: a use case in an international organization. Translating and the Computer 43. Tradulex, (pp 55). <https://archive-ouverte.unige.ch/unige:163351>
- Mees, I. M., Dragsted, B., & Jakobsen, A. L. (2013). Sound effects in translation. *Target*, 25(1), 140–154.
- Mielcarek, M. (2017). *Das simultane Konsektivdolmetschen. (Master's thesis)*. University of Vienna.
- Montecchio, M. (2021). *Defining maximum acceptable latency of ASR-enhanced CAI tools: Quantitative and qualitative assessment of the impact of ASR latency on interpreters' performance (Unpublished MA Thesis)*. Johannes Gutenberg-Universität Mainz.
- Nguyen, H., Estève, Y., & Besacier, L. (2021). An empirical study of end-to-end simultaneous speech translation decoding strategies. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7528–7532). IEEE.
- Niska, H. (2005). Training interpreters: Programmes, curricula, practices. In M. Tennent (Ed.), *Training for the new millennium: Pedagogies for translation and interpreting* (pp. 35–64). John Benjamins.
- Nozaki, J., Kawahara, T., Ishizuka, K., & Hashimoto, T. (2022). End-to-end Speech-to-Punctuated-Text Recognition. *arXiv preprint*. arXiv:2207.03169.
- Orlando, M. (2013, April 8). Interpreting training and digital pen technology. AIIC.net. https://aiic.org/document/885/AIICWebzine_2013_Issue62_8_ORLANDO_Interpreting_training_and_digital_pen_technology_EN.pdf
- Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation and Interpreting*, 6(2), 39–54.

- Orlando, M. (2015a). Digital pen technology and interpreter training, practice and research: Status and trends. In S. Ehrlich & J. Napier (Eds.), *Interpreter Education in the Digital Age* (pp. 125-152). Washington, DC: Gallaudet University Press.
- Orlando, M. (2015b). Implementing digital pen technology in the consecutive interpreting classroom. In D. Andres & M. Behr (Eds.), *To know how to suggest... approaches to teaching conference interpreting* (pp. 171-199). Frank & Timme.
- Ortiz, L. E. S., & Cavallo, P. (2018). Computer-assisted interpreting tools (CAI) and options for automation with automatic speech recognition. *TradTerm*, 32, 9–31. <https://doi.org/10.11606/issn.2317-9511.v32i0p9-31>
- Paone, M. (2016). *Mobile Geräte beim Simultandolmetschen mit besonderem Bezug auf Tablets (Unpublished MA thesis)*. University of Vienna, Austria.
- Pöchhacker, F. (2004). *Introducing interpreting studies*. Routledge.
- Pöchhacker, F. (2016). *Introducing interpreting studies (2nd ed.)*. Routledge.
- Prandi, B. (2015). *The use of CAI tools in interpreters' training: a pilot study*. In *Proceedings of the Translating and the Computer 37 Conference*, London.
- Prandi, B. (2023). *Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology*. Language Science Press.
- Roditi, E. (1982). *Interpreting: Its history in a nutshell*. National Resource Center for Translation and Interpretation. Georgetown University.
- Rodriguez, S. Gretter, R., Matassoni, M., Falavigna, D., Alonso, A., Corcho, O., Rico, M., (2021). SmarTerp: A CAI System to Support Simultaneous Interpreters in Real-Time. In *Proceedings of Triton 2021* (pp. 102-109).
- Roland, R. A. (1982). *Translating world affairs*. Jefferson.
- Rozan, J. (1956). *La prise de notes en interprétation consécutive*. Geneva: Librairie del'Université Georg.
- Russell, D. L., Shaw, R., & Malcolm, K. (2010). Effective teaching strategies for consecutive interpreting. *International Journal of Interpreter Education*, 2, 111-119.
- Russo, M. (2011). Aptitude testing over the years. *Interpreting*, 13(1), 5–30.
- Rütten, A. (2016). Professional precariat or digital elite? Workshop on interpreters' workflows and fees in the digital era. In *Proceedings of Translating and the Computer 38*. London: AsLing.
- Saboo, A., & Baumann, T. (2019). Integration of Dubbing Constraints into Machine Translation. In *Conference on Machine Translation (WMT)* (pp. 94-101). Florence, Italy.

- Seleskovitch, D. (1962). L'interprétation de conférence [Conference interpreting]. *Babel*, 8(1), 13-18.
- Seleskovitch, D., & Lederer, M. (1989). *Pédagogie raisonnée de l'interprétation [Rational pedagogy of interpretation]*. Didier Érudition/OPOCE.
- Seleskovitch, D., & Lederer, M. (1989). *Pédagogie raisonnée de l'interprétation*. Brussels-Luxembourg: OPOCE/Didier Erudition.
- Setton, R. (2015). Simultaneous with text. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 385–386). Routledge.
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., & Wu, Y. (2017). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. *arXiv preprint*. arXiv:1712.05884.
- Shreve, M., Lacruz, I., & Angelone, E. (2010). Cognitive effort, syntactic disruption and visual interference in a sight translation task. In Shreve, M., & Angelone, E. (Eds.), *Translation and Cognition* (pp. 63–84). John Benjamins.
- Singh, N., Khan, R. A., & Shree, R. (2012). Applications of speaker recognition. *Procedia Engineering*, 38, 3122-3126.
- Sperber, M., & Paulik, M. (2020). Speech translation and the end-to-end promise: Taking stock of where we are. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 7409-7421). Virtual Event.
- Stentiford, F. W. M., & Steer, M. G. (1988). Machine Translation of Speech. *British Telecom Technology Journal*, 6(2), 116-122.
- Taylor-Bouladon, V. (2011). *Conference interpreting: Principles and practice*. BookSurge.
- Thiéry, C. (1981). L'enseignement de la prise de notes en interprétation consécutive: un faux problème?. In J. Delisle (Ed.), *L'enseignement de l'interprétation et de la traduction – de la théorie à la pédagogie* (pp. 99–112). Cahiers de traductologie 4, Editions de l'Université d'Ottawa.
- Tommola, J., & Heleva, M. (1998). Language direction and source text complexity effects on trainee performance in simultaneous interpreting. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in diversity* (pp. 177-186). St. Jerome.
- Van Cauwenberghe, G. (2020). *La reconnaissance automatique de la parole en interprétation simultanée*. (MA thesis). Gent University. <https://lib.ugent.be/catalog/rug01:002862551>
- Wadehra, S., Malhotra, S., Tandon, S., & Jain, P. (2021). Comparative Analysis Of Different Speaker Recognition Algorithms. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1-7). <http://doi.org/10.1109/CONIT51480.2021.9498528>.

- Wang, X., & Wang, C. (2018). Can computer-assisted interpreting tools assist interpreting?. *Transletters: International Journal of Translation and Interpreting*, 3, 109-139.
- Weiss, R. J., Chorowski, J., Jaitly, N., Wu, Y., & Chen, Z. (2017). Sequence-to-Sequence Models Can Directly Translate Foreign Speech. In *Proceedings of Interspeech 2017* (pp. 2625-2629). Stockholm, Sweden.
- Weiss, R. J., Chorowski, J., Jaitly, N., Wu, Y., & Chen, Z. (2017). Sequence-to-Sequence Models Can Directly Transcribe Foreign Speech. In *Annual Conference of the International Speech Communication Association (InterSpeech)*, Stockholm, Sweden.
- Will, M. (2020). Computer Aided Interpreting (CAI) for conference interpreters. Concepts, content and prospects. *ESSACHESS-Journal for Communication Studies*, 13(25), 37-71.
- Wiotte-Franz, C. (2001). *Hermeneus und Interpres: zum Dolmetscherwesen in der Antike*. Saarbrücken, Germany: Saarbrücker Druckerei und Verlag.
- Zhang, R., He, Z., Wu, H., & Wang, H. (2022). Learning Adaptive Segmentation Policy for End-to-End Simultaneous Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 7862–7874). Association for Computational Linguistics.

APPENDIX 1. SPEECH MATERIALS

VIOLENCE AGAINST WOMEN /1 (SPEECH B1)

Ladies and gentlemen, The European Agency for Fundamental Rights has just published a study about violence against women. The report is based on interviews with 42,000 women across the 28 Member States of the European Union.

I found the results of this study rather interesting. First of all, there are some horrendous figures. One in three women in Europe has been the victim of violence in adulthood. That means after the age of 15.

One in 20 women has been raped and 75% of women have experienced sexual harassment at work. These are terrible figures...

What I found the most striking about this study was a seeming paradox. In Nordic countries, the risk of violence is greater than in southern countries. This is very bizarre, I think. Because typically we see Scandinavian countries as more egalitarian societies. They're not so male-dominated. Equal opportunities have made great advances and the country's policies reflect this.

For example, they have very good childcare provisions in place so that women can return to work after having children. Or, for example, if we look at the Swedish cabinet, more than half of the cabinet is made up of women and women are very well represented in politics.

But if you look at the figures of this study, in Denmark, 52% of women have experienced violence. In Finland, this rate is 47%, whereas in Spain it's 20%.

If we take a look at all the figures for sexual harassment, 37% of Danish women have experienced sexual harassment compared to 11% in Romania. I found this result very counterintuitive. I would have thought that violence against women was more prevalent in societies that are very patriarchal and where women's roles and women's opinions are secondary to those of a man. I thought violence against women is more in societies where women are often in a submissive role or perhaps even oppressed.

So, when I tried to think about the reasons behind this strange result. The first and depressing question that I asked was Does women's success breed resentment? Is this some sort of horrible backlash in societies where women have become emancipated? And do men simply feel emasculated and does the media portrayal of successful women in those societies substantiate this to some extent? Is it because women are often portrayed as aggressive and power-hungry?

But the study itself suggests a different explanation for these results in Nordic countries. The study says that the emancipation of women in these countries and equal opportunities actually increase the exposure to risk. The authors mean that women work outside the home more in these societies, and they're more likely to go out to meet people. And all of this simply increases your exposure to the risk of violence. Another unconnected factor is urbanization. Countries that are more urbanized see more violence against women.

So, in conclusion, I would say the depressing reality is that whatever the figures, they are much too high. And the real question, to my mind, is what can be done to stop it. Thank you.

VIOLENCE AGAINST WOMEN /2 (SPEECH B2)

Ladies and gentlemen.

Violence is prevalent across all nations and violence against women continues to be a serious social problem worldwide. Violence comes in different forms like physical violence, psychological violence, sexual violence, and economic violence.

In my previous speech, I talked about why the violence against women is more in Scandinavian countries. Now, I would like to touch upon the situation of violence against women in Turkey.

In Turkey, women's rights groups and independent media regularly record hundreds of femicides every year. The number of women killed in Turkey between 2002 and 2015 is 5406.

There are other shocking figures reported by a national study in 2014.

4 of 10 women in Turkey are exposed to physical or sexual violence. 3 of 10 women in Turkey are married before they turn 18.

11% of women are prevented from working by their families. 90% of human trafficking victims in Turkey are women.

The Covid-19 pandemic also had a negative impact on the frequency of domestic violence in Turkey. In poorer provinces, the statistics are more saddening.

But, why is gender-based violence such a problem in Turkey?

First and foremost, patriarchal beliefs are considered a reason why Turkey has a high occurrence of domestic violence. Honour killing is still common in Turkey.

The other reason is economy related. Economic violence against women is a form of violence where women have no financial autonomy. Turkey has the lowest participation of women in the labour force among OECD countries. Economic dependence makes it more difficult for women to leave abusive relationships. A woman dependent on a male family member or partner is more susceptible to other forms of abuse from that person.

When violence takes place within a home, the issue is usually considered a private family matter and becomes ordinary. If we accept violence, if violence receives implicit approval from various social groups, it becomes internalized.

Turkey withdrew from the Council of Europe's convention known as the Istanbul Convention. The convention was designed to combat domestic violence. Turkey was the first country to sign the convention in 2011 and applied in 2014. Withdrawing from the convention means rejecting the international legal norms on gender equality.

To solve the issue of violence against women, we need a holistic approach, from the application of laws to achieving change in mentality and providing education at an early age. And I hope we can do it. Thank you.

EARTHQUAKES IN JAPAN (SPEECH A1)

I would like to express my gratitude to the organizing committee for inviting me here to talk about one of the biggest earthquakes that occurred in Japan's history and the preparedness activities for earthquakes.

Tokyo, one of the most popular and influential cities in Asia, is the capital of Japan, and it is located in the Kanto region on the central Pacific coast of Japan's main island.

Japan's main island is located in an area where four of the Earth's tectonic plates converge. The country is also home to about 10% of the world's active volcanoes.

This means that Japan experiences more earthquakes than anywhere else. Japan experiences around 1,500 earthquakes per year. Japan's long list of earthquakes dates back over a thousand years. Also, when earthquakes occur below or near the ocean, they may trigger tidal waves namely tsunamis. For decades, Japanese people are nervily waiting for a huge earthquake which they called "the big one".

On the 11th of March 2011, many people in Tokyo were hit by a huge quake, swaying wildly. This earthquake was the biggest that ever occurred in Japan's recorded history. It was measured at 9 on the Richter scale. But that was not the expected big earthquake. Although the earthquake in Japan in March last year was extremely big, the epicentre of that quake was not in Tokyo, but 100 kilometres further northeast, in the seas off the coast of Japan.

Tokyo itself was left relatively unharmed by the earthquake last year. but the size of that earthquake and the devastation that it caused in the north of Japan was a reminder to Tokyo that Tokyo needs to be ready for a similar disaster.

It's now estimated that over 20,000 people died in the earthquake, and over quarter of a million people were left sheltering in refuge shelters.

But of course, Tokyo is much more densely populated than the northeast of Japan. The population of Greater Tokyo is about 35 million people. That's a quarter of the total Japanese population. So if a quake hit with its epicentre in or around the Japanese capital, it could be truly devastating.

In terms of the economic costs, according to a study– such an earthquake in or around Tokyo could cause 112 trillion Japanese yen. That's eight times greater than the economic cost of last year's earthquake.

The Research Institute in Japan says that the probability of a serious earthquake in Tokyo within the next 30 years is over 70%.

So the question we have to ask is "is Tokyo ready?"

The answer is perhaps not quite ready.

The Japanese authorities say that all preparation works will be completed soon. So I suppose the residents of Tokyo can only hope and pray that such an earthquake does not arrive before then.

8 WAYS JAPAN IS PREPARED FOR EARTHQUAKES (SPEECH A2)

Ladies and gentlemen,

In my previous speech, I talked to you about the biggest earthquake recorded in Japan's history and how Tokyo can be affected by a similar or bigger earthquake.

As Japan experience earthquakes regularly, they've become one of the best-prepared nations on earth. The ability to innovate, invest, educate, and learn from past mistakes has made Japan the most earthquake-ready country in the world.

Now I would like to talk about eight ways Japan prepares for earthquakes. Firstly, Japan builds earthquake-resistant buildings. All houses are built to resist some level of tremor. Houses in Japan are built to comply with rigorous earthquake-proof standards that have been set by law. These laws also apply to other structures like schools and office buildings. It's said that around 87% of the buildings in Tokyo can resist earthquakes.

Second of all, phone updates. Every smartphone in Japan is installed with an earthquake and tsunami emergency alert system. Before the impending disasters, the alarm is triggered for around five to ten seconds. The alert system give users time to quickly seek protection.

Thirdly, Japan's high-speed trains are equipped with earthquake sensors that are triggered to freeze every moving train in the country if necessary. In 2011, when a 9.0 magnitude quake hit Japan, there were 27 trains in action.

Fourthly, if an earthquake hits the nation, all of Japan's TV channels immediately switch to official earthquake coverage, ensuring that the population is well informed on how to stay safe.

As a fifth measure, schools in Japan run regular earthquake drills once a month. From a young age, schoolchildren are educated on the best way to seek protection and stay safe if an earthquake hits their area. Another way Japan helps protect its population against future natural disasters is by learning from past events. In 1995, the city of Kobe was struck by a completely devastating earthquake, which killed 5,000 people and destroyed tens of thousands of homes. After the city is rebuilt, the Kobe Earthquake Memorial Museum was constructed in the city.

The next thing that is important for Japan's preparation is the earthquake survival kits. Every household has to keep a survival kit with a flashlight, a radio, a first aid kit and enough food and water to last for a few days.

Lastly, the water discharge tunnel is one of the most impressive feats of Japanese engineering. This large and hidden tunnel collects flood waters caused by natural disasters like cyclones and tsunamis and safely redistributes the water into the Edo River. It took 13 years to build this massive tunnel and it cost 3 billion dollars, but you cannot put a price on how many lives it promises to save.

Thank you,

APPENDIX 2. TABLE OF ICT TOOLS AND PLATFORMS RELATED TO INTERPRETING TECHNOLOGY

Name	Category (main function)	SPECIFICITY	PURPOSE			MODALITY		FEATURE			
			Interpreter Training	Prep. (Corpora Building)	Prep. (Terminology Management)	Simultaneous	Consecutive	RI Platform	ASR	(R) Advanced Booth Controls	Replacement
Meliss KOSMOS	Training Platform	Y	X			X	X				
InTrain	Training Platform	Y	X			X	X				
Linkinterpreting	Training Platform	Y	X			X	X				
Interpreter Training Resources.eu	Training Platform	X	X			X	X				
InterpreterQ Media Player	Training Platform	Y	X								
Speechpool	Speech Bank	Y	X								
ORCIT	Speech Bank	Y	X								
EU DG -SCIC	Speech Bank	Y	X								
Speech Rep.											
Interplex UE	Glossary Management	Y			X	X					
VIP Voice-text Integrated System for Interpreters	Glossary Management	Y	X		X	X	X		X		
InterpreBank	Glossary Management	Y			X	X	X				
KUDO Interpreter Assist	Glossary Management	Y			X	X	X		X		
Interpreter's Help	Glossary Management	Y			X	X	X				
Flashterm	Glossary Management	N			X						
Intragloss	Glossary Management	Y			X		X				
BooCAT	Corpora Building	Y		X							
SDL Multiterm Extract	Terminology Extraction	Y			X						
Simple Extractor	Terminology Extraction	Y			X						
Sketch Engine	Terminology Extraction	Y			X						
Terminus	Terminology Extraction	Y			X						
TermSuite	Terminology Extraction	Y			X						
InterpreBank ASR	Speech Recognition	Y				X	X		X		
Dragon NS	Speech Recognition	N				X	X				
Evernote	Note-taking	N						X			
Cymo Note	Note-taking	Y						X			
Slight-T-ep	Note-taking	Y						X			
Neo SmartPen	Note-taking	N									
Livescribe Smart Pen	Note-taking	N						X			
Nebo	Note-taking	N						X			
Bamboo Paper	Note-taking	N						X			
Noteshelf	Note-taking	N						X			
Notability	Note-taking	N						X			
Penultimate	Note-taking	N						X			
LectureNotes	Note-taking	N						X			
CymoBooth	Virtual Booth Service	Y							X	X	
Smart-T-ep	Audio and Video Conference	Y							X	X	
GreenT-ep	Audio and Video Conference	N	X						X	X	
KUDO	Audio and Video Conference	N							X	X	
Converso	Audio and Video Conference	N							X	X	
Olyset	Audio and Video Conference	N							X	X	
Interactio	Audio and Video Conference	N							X	X	
eAPiPico	Audio and Video Conference	N							X	X	
VoiceBoxer	Audio and Video Conference	N							X	X	
InterpreTy	Audio and Video Conference	N							X	X	
QuaQua	Audio and Video Conference	N							X	X	
Akkadu	Audio and Video Conference	N							X		
Catalava	Audio and Video Conference	N							X		
CymoMeeting	Audio and Video Conference	N							X		
Lingolet	Audio and Video Conference	N							X		
Ablicoconference	Audio and Video Conference	N							X		
InterpreCloud	Audio and Video Conference	N							X		
TranslIRSI	Audio and Video Conference	N							X		
Wordsynk	Audio and Video Conference	N							X		

APPENDIX 3. ETHICS COMMITTEE APPROVAL



T.C.
HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜ
Rektörlük

Tarih: 17/04/2023 09:38
Sayı: E-35853172-300-00002802234



Sayı : E-35853172-300-00002802234
Konu : Etik Komisyon İzni (Cihan ÜNLÜ)

17.04.2023

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 10.04.2023 tarihli ve E-12908312-300-00002790892 sayılı yazınız.

Enstitünüz Mütercim Tercümanlık Anabilim Dalı İngilizce Mütercim Tercümanlık Yüksek Lisans Programı öğrencilerinden **Cihan ÜNLÜ**'nün, **Prof. Dr. Aymil DOĞAN** danışmanlığında hazırladığı **“Otomatik Konuşma Tanıma Sistemlerinin Ardıl Çeviride Kullanılması: Sight-Terp”** başlıklı tez çalışması Üniversitemiz Senatosu Etik Komisyonunun **11 Nisan 2023** tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Sibel AKSU YILDIRIM
Rektör Yardımcısı

Bu belge güvenli elektronik imza ile imzalanmıştır.

Belge Doğrulama Kodu: C5477560-F7FF-410D-ACC3-1C808714FFEF

Belge Doğrulama Adresi: <https://www.turkiye.gov.tr/hu-ebys>

Adres: Hacettepe Üniversitesi Rektörlük 06100 Sıhhiye-Ankara

Bilgi için: Çağla Handan GÜL

E-posta: yaziml@hacettepe.edu.tr İnternet Adresi: www.hacettepe.edu.tr Elektronik

Bilgisayar İşletmeni

Ağ: www.hacettepe.edu.tr

Telefon: 03123051008

Telefon: 0 (312) 305 3001-3002 Faks: 0 (312) 311 9992

Kep: hacettepeuniversitesi@hs01.kep.tr



APPENDIX 4. THESIS/DISSERTATION ORIGINALITY REPORT



**HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
MASTER'S THESIS ORIGINALITY REPORT**

**HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
DEPARTMENT OF TRANSLATION AND INTERPRETING**

Date: 03/05/2023

Thesis Title : AUTOMATIC SPEECH RECOGNITION IN CONSECUTIVE INTERPRETER WORKSTATION:
COMPUTER-AIDED INTERPRETING TOOL 'SIGHT-TERP'

According to the originality report obtained by myself/my thesis advisor by using the Turnitin plagiarism detection software and by applying the filtering options checked below on 04/05/2023 for the total of 98 pages including the a) Title Page, b) Introduction, c) Main Chapters, and d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 10 %.

Filtering options applied:

1. ☒ Approval and Declaration sections excluded
2. ☒ Bibliography/Works Cited excluded
3. ☒ Quotes excluded
4. ☐ Quotes included
5. ☒ Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Social Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

03.05.2023

Name Surname: CİHAN ÜNLÜ

Student No: N20137604

Department: MÜTERCİM VE TERCÜMANLIK

Program: İNGİLİZCE MÜTERCİM VE TERCÜMANLIK

ADVISOR APPROVAL

APPROVED.

PROF. DR. AYMİL DOĞAN

(Title, Name Surname, Signature)