

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS
LINGUÍSTICOS

FLÁVIO DE SOUSA FREITAS

**INTERPRETERS IN THE MACHINE:
EXPLORING HUMAN ROLE IN THE
DEVELOPMENT OF SPEECH-TO-SPEECH
TRANSLATION SYSTEMS**

Uberlândia

2026

FLÁVIO DE SOUSA FREITAS

**INTERPRETERS IN THE MACHINE:
EXPLORING HUMAN ROLE IN THE
DEVELOPMENT OF SPEECH-TO-SPEECH
TRANSLATION SYSTEMS**

PhD dissertation submitted to the Programa de Pós-Graduação em Estudos Linguísticos of the Instituto de Letras e Linguística of the Universidade Federal de Uberlândia as a partial requirement for the degree of PhD in Applied Linguistics.

Field of studies: Linguistics and Applied Linguistics.

Research line: Language, teaching, and society

Academic Advisor: Prof. Dr. Marileide Dias Esqueda.

Uberlândia

2026

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

F866 2026	Freitas, Flávio de Sousa, 1992- INTERPRETERS IN THE MACHINE [recurso eletrônico] : EXPLORING HUMAN ROLE IN THE DEVELOPMENT OF SPEECH-TO- SPEECH TRANSLATION SYSTEM / Flávio de Sousa Freitas. - 2026. Orientadora: Marileide Dias Esqueda. Tese (Doutorado) - Universidade Federal de Uberlândia, Pós- graduação em Estudos Linguísticos. Modo de acesso: Internet. DOI http://doi.org/10.14393/ufu.te.2026.507 Inclui bibliografia. 1. Linguística. I. Esqueda, Marileide Dias, 1973-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Estudos Linguísticos. III. Título. CDU: 801
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Estudos Linguísticos

Av. João Naves de Ávila, nº 2121, Bloco 1G, Sala 1G256 -
Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902

Telefone: (34) 3239-4102/4355 - www.ileel.ufu.br/ppgel -

secpiegel@ileel.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Estudos Linguísticos				
Defesa de:	Tese de Doutorado - PPGEI				
Data:	Sete de abril de dois mil e vinte e seis	Hora de início:	09:00	Hora de encerramento:	12:00
Matrícula do Discente:	12213ELI008				
Nome do Discente:	Flávio de Sousa Freitas				
Título do Trabalho:	INTERPRETERS IN THE MACHINE: EXPLORING HUMAN ROLE IN THE DEVELOPMENT OF SPEECH-TO-SPEECH TRANSLATION SYSTEMS				
Área de concentração:	Estudos em Linguística e Linguística Aplicada				
Linha de pesquisa:	Linguagem, Ensino e Sociedade				
Projeto de Pesquisa de vinculação:	Ensino e Aprendizagem de Tradução e Interpretação: pressupostos teóricos, epistemológicos e metodológicos				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Estudos Linguísticos, assim composta: Professores Doutores: Ariel Novodvorski - UFU; Helena de Almeida Maia - UNICAMP; Daniel Antônio de Sousa Alves - UFPB; Glória Regina Loreto Sampaio - PUC (SP); e Marileide Dias Esqueda (UFU), orientadora do candidato.

Iniciando os trabalhos, a presidente da mesa, Dra. Marileide Dias Esqueda, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir, a senhora presidente concedeu a palavra, pela ordem sucessivamente,

aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após o cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar, os trabalhos foram encerrados. Em seguida, foi lavrada a presente ata que, após lida e achada conforme, foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Marileide Dias Esqueda, Professor(a) do Magistério Superior**, em 05/05/2026, às 14:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Helena de Almeida Maia, Usuário Externo**, em 05/05/2026, às 15:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Glória Regina Loreto Sampaio, Usuário Externo**, em 05/05/2026, às 19:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Daniel Antonio de Sousa Alves, Usuário Externo**, em 06/05/2026, às 06:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ariel Novodvorski, Professor(a) do Magistério Superior**, em 06/05/2026, às 11:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7283608** e o código CRC **C4FC28D1**.

To my family and to all those who seek
knowledge and believe in the
transformative power of education.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Marileide Dias Esqueda, for her guidance, encouragement, and invaluable support throughout the development of this doctoral research. Her knowledge, patience, and commitment were fundamental to the completion of this work.

I am also sincerely grateful to the members of the examination committee for their time, careful reading, and valuable comments and suggestions, which have greatly contributed to the improvement of this thesis.

My heartfelt thanks go to GETTEC research group, to the PPGEL, and to the Federal University of Uberlândia for providing the academic environment, support, and opportunities that made this research possible.

I would also like to acknowledge CAPES for the scholarship granted during my doctoral studies, which was essential for the development of this research.

I am deeply thankful to my family for their love, support, and constant encouragement throughout this journey. My special thanks go to my father, Divino Juvencio de Freitas, my mother, Albina Rosa de Sousa, my sisters, Bruna and Marina, and my brother, Daniel. Their presence and support have always been a source of strength to me.

I would like to express my loving gratitude to my boyfriend, Welberth, for his companionship, understanding, encouragement, and support during the many challenges of this doctoral journey.

Finally, I would like to thank all the people who, in one way or another, helped me throughout my doctorate. Whether through academic support, friendship, encouragement, or kindness, each contribution was meaningful and will always be remembered.

“Technology is usually fairly neutral. It’s like a hammer, which can be used to build a house or to destroy someone’s home. The hammer doesn’t care. It’s almost always up to us to determine whether the technology is good or bad.”

– Noam Chomsky

ABSTRACT

This dissertation examines the role of professional interpreters (PIs) in the field of Speech-to-Speech Translation (S2ST), moving beyond the conventional assumption that their participation is largely restricted to data collection, annotation, and system evaluation. Although S2ST has progressed substantially as a technological response to multilingual communication needs, the broader potential of interpreter competence remains insufficiently explored in the literature. In light of this gap, the present study investigates how the knowledge and skills that constitute PI competence may contribute to a wider range of stages in the development of S2ST systems, while also reconsidering the relationship between human interpreters and language technologies from a complementary, rather than substitution-oriented, perspective. Methodologically, the research adopts a mixed-methods design based on the compilation of two domain-specific corpora of open-access academic publications retrieved from Google Scholar through Publish or Perish. The first corpus, focused on Interpreting, comprises 264 publications, while the second, dedicated to S2ST, contains 953 publications. After retrieval, metadata cleaning, selection, and file organization, both corpora were submitted to semantic and contextual analysis using ChatGPT-o1. The model was employed to identify, on the one hand, the principal domains of PI competence and, on the other, the main development stages of S2ST systems. The resulting reports were then comparatively examined in order to determine the stages at which interpreters' expertise could make meaningful contributions. Given the known limitations of generative AI, particularly the risk of hallucination, the analyses were constrained to the uploaded corpora and subsequently subjected to manual verification against the source materials. The findings indicate that PI competence is a multidimensional construct encompassing linguistic and communicative mastery, cultural and contextual sensitivity, cognitive agility, discourse management, terminology handling, and ethical-professional judgment. Likewise, the study identifies seven major stages in the development of S2ST systems and demonstrates that the competence of professional interpreters may be relevant across far more stages than is commonly acknowledged. Rather than functioning solely as providers of training data or as evaluators of output quality, interpreters emerge as potential contributors to system design, communicative modeling, interface refinement, prosodic and pragmatic calibration, bias mitigation, and ethical risk assessment. This dissertation argues, therefore, that S2ST development can benefit from a

more systematic integration of interpreter expertise. By proposing a framework that positions PIs as collaborators in the design of more human-centered, inclusive, and ethically responsive technologies, the study contributes to ongoing discussions at the intersection of Interpreting Studies, artificial intelligence, and multilingual communication.

Keywords: professional interpreter, speech-to-speech translation, interpreting studies, artificial intelligence

RESUMO

Esta tese examina o papel dos intérpretes profissionais (IPs) no campo da interpretação automática (S2ST, do inglês speech-to-speech translation), indo além da suposição convencional de que sua participação se restringe, em grande medida, à coleta de dados, à anotação e à avaliação de sistemas. Embora a tecnologia S2ST tenha avançado substancialmente como resposta tecnológica às necessidades de comunicação multilíngue, o potencial mais amplo da competência do intérprete permanece ainda pouco explorado na literatura. Diante dessa lacuna, o presente estudo investiga de que maneira os conhecimentos e as habilidades que constituem a competência do IP podem contribuir para uma gama mais ampla de etapas no desenvolvimento de sistemas S2ST, ao mesmo tempo em que reconsidera a relação entre intérpretes humanos e tecnologias da linguagem sob uma perspectiva de complementaridade, e não de substituição. Metodologicamente, a pesquisa adota uma abordagem de métodos mistos com base na compilação de dois corpora, compostos por publicações acadêmicas de acesso aberto recuperadas no Google Scholar por meio do software Publish or Perish. O primeiro corpus, centrado em Interpretação, é composto por 264 publicações, enquanto o segundo, dedicado à S2ST, contém 953 publicações. Após a recuperação, a limpeza dos metadados, a seleção e a organização dos arquivos, ambos os corpora foram submetidos à análise semântica e contextual com o uso do ChatGPT-o1. O modelo foi empregado para identificar, por um lado, os principais domínios da competência do IP e, por outro, as principais etapas de desenvolvimento dos sistemas de S2ST. Os relatórios resultantes foram então examinados comparativamente com o objetivo de determinar em quais etapas a expertise dos intérpretes poderia oferecer contribuições significativas. Considerando as limitações conhecidas da inteligência artificial generativa, especialmente o risco de alucinação, as análises foram restringidas aos corpora anexados ao ChatGPT-o1 e, posteriormente, submetidas à verificação manual com base nos materiais-fonte. Os resultados indicam que a competência do IP é um fenômeno que abrange domínio linguístico e comunicativo, sensibilidade cultural e contextual, agilidade cognitiva, gerenciamento discursivo, manejo terminológico e julgamento ético-profissional. Do mesmo modo, o estudo identifica sete etapas principais no desenvolvimento de sistemas S2ST e demonstra que a competência dos IPs pode ser relevante em muito mais etapas do que se costuma reconhecer. Em vez de atuarem apenas como fornecedores de dados de treinamento ou avaliadores da qualidade das saídas do sistema, os intérpretes emergem como potenciais colaboradores no

design de sistemas, na modelagem comunicativa, no refinamento de interfaces, na calibração prosódica e pragmática, na mitigação de vieses e na avaliação de riscos éticos. Esta tese sustenta, portanto, que o desenvolvimento da tecnologia S2ST pode se beneficiar de uma integração mais sistemática da expertise dos intérpretes. Ao propor um quadro teórico que posiciona os IPs como colaboradores no desenvolvimento de tecnologias mais centradas no humano, inclusivas e eticamente responsáveis, o estudo contribui para as discussões em curso na interseção entre os Estudos da Interpretação, a inteligência artificial e a comunicação multilíngue.

Palavras-chave: intérprete profissional, speech-to-speech translation, estudos da interpretação, inteligência artificial

LIST OF TABLES

Table 1 - ChatGPT request for more information.	43
Table 2 - Structure of the report generated by ChatGPT.	44
Table 3 - Interpreting corpus size.	64
Table 4 - Linguistic and communicative competence.	65
Table 5 - Cognitive and memory skills.	66
Table 6 - Cultural and contextual awareness.	67
Table 7 - Subject-matter knowledge.	68
Table 8 - Ethical and professional conduct.	69
Table 9 - S2ST Corpus size.	71
Table 10 - Data collection & Preparation.	71
Table 11 - ASR model development.	72
Table 12 - Text translation model development.	73
Table 13 - TTS model development.	74
Table 14 - Integration and real-time system design.	76
Table 15 - Evaluation and iteration.	77
Table 16 - Maintenance and continuous improvement.	78
Table 17 - Tensions between the ideal and the technical reality of S2ST systems	94
Table 18 - The "Mediator" Dilemma.	96
Table 19 - Ethical Ideals vs. Technical Constraints	99
Table 20 - The PI as the gold standard.	101
Table 21 - Download automation of the academic publications on S2ST.	136

LIST OF FIGURES

Figure 1 - Keywords and period.	22
Figure 2 - RIS file.	23
Figure 3 - Retrieved references information and metadata.	23
Figure 4 - Corpora compilation and organization.	25
Figure 5 - Retrieved publications attached, advanced reasoning model.	26
Figure 6 - Report generated by ChatGPT.	27
Figure 7 - ChatGPT research activity.	40

LIST OF ABBREVIATIONS

AI – Artificial Intelligence

AIIC – International Association of Conference Professional Interpreters

API – Application Programming Interface

ASR – Automatic Speech Recognition

AUSIT – Australian Institute of Interpreters and Translators

AXCell – Automatic Extraction of Results from Machine Learning Papers

BCE – Before Common Era

BLEU – Bilingual Evaluation Understudy

BPE – Byte Pair Encoding

CEO – Chief Executive Officer

CTC – Connectionist Temporal Classification

DNNs – Deep Neural Networks

DOI – Digital Object Identifier

E2E – End-to-end

ESIT – Ecole Supérieure d’Interprètes et de Traducteurs

EBSCO – Elton B. Stephens Company

GMMs – Gaussian Mixture Models

GPT – Generative Pre-trained Transformer

GS – Google Scholar

GSTs – Global Style Tokens

HMMs – Hidden Markov Models

HTML – HyperText Markup Language

HTTP – HyperText Transfer Protocol

IBM – International Business Machines

IS – Interpreting Studies

JSON – JavaScript Object Notation

kNN – K-Nearest-Neighbor

MB – Mega Bytes

METEOR – Metric for Evaluation of Translation with Explicit Ordering

MFCCs – Mel-frequency Cepstral Coefficients

MOS – Mean Opinion Score

MT – Machine Translation

NAJIT – National Association of Judiciary Interpreters and Translators

NCIHC – National Council on Interpreting in Health Care

NLP – Natural Language Processing

NMT – Neural Machine Translation

NNs – Neural Networks

NTTS – Neural Text-to-Speech

PDF – Portable Document Format

PhD – Philosophiae Doctor

PI - Professional Interpreter

PoP – Publish or Perish

PPGEL – Programa de Pós-graduação em Estudos Linguísticos

RID – Registry of Interpreters for the Deaf

RIS – Research Information Systems

RNNs – Recurrent Neural Networks

ROUGE – Recall-Oriented Understudy for Gisting Evaluation

Seq2Seq – Sequence-to-sequence

SerpAPI – Search Engine Results Page Application Programming Interface

S2ST – Speech-to-Speech Translation

TIS – Translation and Interpreting Studies

TS – Translation Studies

TTS – Text-to-Speech

UI – User Interface

UN – United Nations

US – United States

UX – User Experience

VR – Virtual Reality

Wav2Vec – Waveform to Vector

WER – Word Error Rate

WRHA – Winnipeg Regional Health Authority

WWI – World War I

Table of Contents

1. Introduction	17
1.1. Motivation	17
1.2. Rationale	18
1.3. Outline	19
1.3.1. Part I	19
1.3.2. Part II	20
2. Research design	22
2.1. Compiling the corpora	22
2.1.1. Retrieving academic publications	23
2.1.2. Metadata editing	24
2.1.3. Downloading the academic publications	25
2.2. Identifying the stages of development of S2ST systems and PIs' competence	26
PART I	29
3. Related work	30
3.1. Interpreting and technology	30
3.2. PIs as the reference standard	34
3.3. Artificial intelligence-based semantic and contextual analysis	35
4. Who is the PI?	40
4.1. Historical evolution of the professional interpreter's role	42
4.2. Social, ethical, and professional identity of interpreters	47
5. Speech-to-speech translation	52
5.1. The Three components	53
5.2. E2E S2ST Models	57
PART II	59
6. Professional interpreter's competence	61
6.1. Linguistic and communicative competence	61
6.2. Cognitive and Memory Skills	63
6.3. Cultural and contextual awareness	64
6.4. Subject-matter competence	65
6.5. Ethical and professional conduct	66
7. Speech-to-speech development stages	67
7.1. Data collection and preparation	67
7.2. ASR model development	68
7.3. MT model development	69
7.4. TTS model development	70
7.5. Integration and real-time system design	72
7.6. Evaluation and iteration	73

7.7. Maintenance and continuous improvement	74
8. Potential contributions of PIs to S2ST	76
8.1. Design and User Experience (UX)	77
8.2. Linguistic and cultural adaptation	79
8.3. Ethical Considerations	82
8.4. Real-world deployment	84
9. Results and Discussion	89
9.1. Discussing Design and User Experience (UX)	89
9.2. Discussing linguistic and cultural adaptation	91
9.3. Discussing ethical Considerations	94
9.4. Discussing real-world deployment	97
REFERENCES	100
APPENDIX I - Script	162
APPENDIX II - Literature Review Generated by ChatGPT	165
1. Historical Evolution of the Professional interpreter's Role	165
2. Social, Ethical, and Professional Identity of Professional interpreters	167

1. Introduction

The first part of this introductory chapter presents the motivation behind this study, outlining its central focus and specific objectives. The second part contextualizes the academic path that led to the formulation of this research, along with the underlying rationale that guided its development.

1.1. Motivation

This dissertation investigates the expanded role of Professional Interpreters (PIs) in Speech-to-Speech Translation (S2ST) development. Moving beyond traditional data annotation, it argues that PI competence is vital across all technical stages. By integrating their unique contextual and cultural awareness, S2ST systems can evolve from mere substitutes to sophisticated, human-informed tools.

S2ST is a technology that converts spoken words from one language into another and plays back the translation. Its main goal is to overcome language barriers and make communication easier.

In crisis situations, humanitarian workers can use S2ST tools to talk directly with displaced populations who speak different languages. This ensures important information is communicated accurately and quickly.

In schools and other educational settings, S2ST helps students access resources in any language, enabling them to participate in important conversations. This can potentially give access to knowledge, no matter what language students speak.

S2ST technology could significantly change the interpreting profession. This shift comes with both challenges and new opportunities.

On the one hand, S2ST systems might decrease the demand for PIs in routine or less critical situations because they provide real-time multilingual communication. On the other hand, given the limitations of S2ST, PIs will still be necessary in situations where their unique knowledge and skills are demanded.

Unlike S2ST systems, PIs can capture irony, sarcasm, cultural allusions, subtext and many other aspects crucial to understanding a spoken or signed message. Despite their superiority over machines, however, PIs have been employed mainly in the stages of data collection and annotation and in the stages of evaluating S2ST systems.

While this view recognizes the importance of PIs as information managers in ensuring the quality and accuracy of training data, it tends to neglect the wider potential of their competence. By confining PIs to data supplying and evaluation of output, S2ST research has failed to explore how the competence of these professionals could improve the practical application of S2ST technologies or how these tools might complement PIs in their profession (Pöchhacker, 2024).

In view of this, what other stages of S2ST development could PIs contribute to? To address such a question, the main focus of this dissertation is to understand the role of PIs within S2ST in addition to data training and output evaluation.

The specific objectives of this dissertation are

- i. To identify the knowledge and skills that constitute the competence of PIs and their influences;
- ii. To envisage in which stages of the development of S2ST systems PIs can perform;
- iii. To call into question the interaction between PIs and S2ST technologies.

Within the evolving landscape of Interpreting Studies, technology is typically categorized into three distinct layers: tools that support the interpreter's delivery (such as terminology banks and specialized datasets), platforms that mediate the service and facilitate training (such as Zoom or RSI platforms), and finally, the realm of automation and Artificial Intelligence. This third category includes machine interpreting applications designed to assist end-users or, in some cases, to function as a substitute for human professionals. In this dissertation, we will focus our discussion on the latter and their relation to PIs.

1.2. Rationale

Our interest in the intersection of interpreting and technology arose approximately 10 years ago, during our undergraduate studies in Translation. In 2016, our discussions on this topic with our advisor professor culminated in the formulation of our final course paper entitled *The State of the Art in Speech-to-Speech Translation: From Post-War to Speech Translation Apps* (Freitas, 2016), which we publish as a book in 2020 (Freitas & Esqueda, 2020).

In 2018, I was admitted to the Master's program in the Programa de Pós-graduação em Estudos Linguísticos (PPGEL), becoming a member of the research line 3, which investigates language, teaching, and society. During that period, our research focused on the expectations and perceptions of Brazilian users regarding S2ST mobile applications (Freitas, 2022).

We are currently affiliated with the same research line and the same graduate program. At this moment, however, we consider that a significant opportunity in the development of S2ST systems is focused on the professional interpreter. Therefore, this dissertation proposes an integrated approach, in which PIs go far beyond traditional tasks such as data supplying and evaluation.

We consider it of utmost importance to clarify that the choice of the term “competence” is due to its widespread use in the relevant literature. In this study, therefore, competence refers to the set of knowledge and skills developed by PIs to perform their tasks effectively.

Although it is beyond the scope of our study, we are aware of the ongoing debates regarding the use of this term. Esqueda (2025), for instance, argues that “due to its standardized nature, oriented towards the job market, economic competitiveness and efficiency, competence-based training leaves little room for the education of creative and independent individuals (...)” (p. 24).

This study notes the shifting dynamics of the interpreting profession (Okoniewska, 2022); at a time in which technology is redefining roles and responsibilities (Behrend et al., 2024). This research stresses the need to involve PIs in shaping future communication tools in a multidisciplinary way, working closely with computer linguists, technologists, and developers.

This study also aims to address a critical gap in current scientific literature: the lack of discussion on the potential contributions by PIs throughout the S2ST lifecycle. Seeking to visualize the potential S2ST research areas PIs can contribute to; this dissertation provides a comprehensive framework for future investigations and practical advancements.

Such a framework benefits both academics, and consequently the industry and society as a whole. It heads toward humanized, inclusive S2ST systems. And with that, the central theme of this dissertation emerges to be PIs' competence not only as the driving force behind technological advances but also in line with complexity and convenience for communication.

1.3. Outline

The following chapter, Chapter 2, describes the methodological procedures through which we conducted this study. All subsequent chapters will be presented in two parts. Part I establishes the theoretical and contextual foundations of the study. Part II performs the analytical synthesis which systematizes professional interpreter competence into clearly defined domains; breaks down S2ST into concrete development stages and maps PI's competence onto stages to demonstrate where, how, and why human expertise can possibly improve S2ST systems.

1.3.1. Part I

Chapter 3 reviews related studies on interpreting and technology, the role of PIs and emphasizing their competence, ethical considerations, and how they serve as performance benchmarks for AI systems. It reviews technological advancements in interpreting, such as remote and hybrid modes, and highlights scholarly perspectives that assert the irreplaceability of human traits like empathy, contextual awareness, and interpretive finesse in high-stakes communication.

Additionally, Chapter 3 analyzes the use of AI, particularly ChatGPT, as a tool for academic research and semantic analysis in S2ST studies. While AI aids in literature synthesis, trend detection, and data organization, its limitations—such as fabricating content or misinterpreting context—necessitate critical human oversight.

Chapter 4 explores the historical development and evolving identity of PIs, emphasizing their transition from informal linguistic mediators in ancient and colonial contexts to formally trained experts in institutional and high-stakes environments. It traces key milestones such as the Paris Peace Conference, the Nuremberg Trials, and the professionalization of the field through specialized training programs and the emergence of Interpreting Studies as an academic discipline.

Chapter 5 provides an overview of S2ST, detailing its three core components, and discussing the evolution, challenges, and advancements of each, including the rise of end-to-end (E2E) models. It emphasizes the interdisciplinary nature of S2ST, technical and ethical challenges, and the ongoing efforts to enhance real-time multilingual communication through neural architectures and sociolinguistic insights.

1.3.2. Part II

Chapter 6 identifies five core domains that constitute the competence of a professional interpreter. Drawing from a curated corpus of 264 academic publications, it outlines the specific abilities interpreters must master—such as idiomatic fluency, multitasking, and domain-specific terminology management—to perform effectively and accurately across various interpreting contexts.

In Chapter 6 data emphasize that professional interpreting requires not only technical and linguistic skills but also deep cultural sensitivity, rapid cognitive processing, and unwavering ethical standards like neutrality and confidentiality. By synthesizing AI-generated insights with academic literature, Chapter 6 demonstrates how interpreter competence is a complex, multidimensional construct essential to high-quality human-mediated communication.

Chapter 7 outlines seven key development stages of a S2ST system. Each stage is described through insights derived from a 953-publication corpus, illustrating how diverse tasks combine to create robust, real-time translation systems.

The chapter emphasizes the interdisciplinary complexity of building S2ST systems, integrating linguistic, computational, and human-centered components. From designing efficient speech pipelines and reducing latency to ensuring evaluation through human feedback and metric-based analysis, it presents a comprehensive blueprint for creating adaptive and scalable S2ST solutions that mirror the nuances of human communication.

Chapter 8 explores potential contributions PIs can make to the development of S2ST systems by enhancing user experience, linguistic quality, cultural sensitivity, and ethical integrity. Drawing on their competence in real-time communication, discourse coherence, and cross-cultural awareness, interpreters can inform error handling, interface design, prosodic modeling, and bias mitigation, ensuring S2ST systems deliver nuanced, respectful, and intelligible translations.

Finally, in Chapter 9, we summarize and discuss our results.

2. Research design

This dissertation employs a qualitative, interdisciplinary methodology, blending Interpreting Studies with Human-Computer Interaction principles. Through a systematic literature review and comparative analysis of S2ST development lifecycles, the research identifies critical integration points for professional expertise. This approach evaluates how interpreting competence can be codified into technical workflows, moving beyond simple data annotation to strategic, human-in-the-loop architectural design.

We compiled two corpora of academic works on interpreting and S2ST. The two corpora include academic publications freely available on Google Scholar (GS)¹. Both corpora were attached to ChatGPT-o1², 1.2025.41 version, which performed semantic analyses to identify the knowledge and skills that make up the competence of PIs and to identify the main stages of development of S2ST systems. We submitted the generated reports to ChatGPT-o1 to identify which S2ST systems development stages PIs can contribute to.

Several studies indicate that conversational AIs can produce hallucinations (Safar et al., 2025; Mills & Angell, 2025; Xiao et al., 2025). They generate information that appears plausible and coherent, but is factually incorrect, fabricated, or not grounded in training data or reality.

However, according to Roozbahani (2025), this behavior can be mitigated by providing examples or parameters upon which the AI can ground its responses. It is as though, rather than querying its entire training database, the AI would concentrate solely on the user-provided data.

Considering this, we compiled two domain-specific corpora and uploaded them to ChatGPT-o1. In this way, we delimit data to be analysed in order to avoid generalizations and reduce hallucinations.

2.1. Compiling the corpora

We carried out the compilation of the corpora for this study in three stages: retrieving academic publications, editing metadata and downloading publication files. The entire

¹ <https://scholar.google.com/>.

² <https://chatgpt.com/>.

compilation process was done with free tools and took 386 days to complete, from March 2023 to April 2024.

2.1.1. Retrieving academic publications

Data in this research were extracted from academic references retrieved through the software Publish or Perish (PoP) (Harzing, 2010), version 8.17.4863.9118. Inside the PoP interface we selected the GS search engine.

For the Interpreting corpus we applied the keywords “interpreting” and “interpretation”. For the S2ST corpus we applied the keywords “speech translation” and “speech-to-speech translation”, in accordance with the findings of Freitas and Esqueda (2017, p.116).

The search included only citations, and due to the technological limitations of PoP, the maximum number of documents selected is 1,000. For both corpora, we defined the period from 2014 to 2024, as shown in Figure 1.

Figure 1 - Keywords and period.

The figure displays two screenshots of the Google Scholar search interface. Both screenshots show the search bar with the keywords 'interpreting AND interpretation' (top) and 'speech translation AND speech-to-speech translation' (bottom). The 'Maximum results' are set to 1000, and the 'Years' range is set from 2014 to 2024. The 'CITATIONS' checkbox is selected in both cases, while 'Patents' and 'Only review articles' are unselected.

Source: PoP interface.

Although PoP allows access to the files of open-access references retrieved through its interface, it is not possible to edit information in case of inconsistencies in the metadata. Therefore, we saved the search results in RIS format (Azeroual et al., 2018), which includes information such as the authorship, title, year of publication, and other relevant details, as shown in Figure 2.

Figure 2 - RIS file.

```

TY  - JOUR
AU  - Smith, John
AU  - Doe, Jane
TI  - Advances in Speech-to-Speech Translation
T2  - Journal of Computational Linguistics
PY  - 2023
DO  - 10.1234/jcl.2023.456789
SN  - 1234-5678
AB  - This study explores the latest advances in speech-to-speech translation systems.
KW  - speech-to-speech translation; machine learning; NLP
UR  - https://journalwebsite.com/article/456789

```

Source: the author.

The process of retrieving the references took about an hour. The computer used in this and the other stages of the study is a Dell Inspiron 15 Laptop, Windows 11 operating system.

2.1.2. Metadata editing

We then uploaded the RIS file to the reference management tool Mendeley (Reiswig, 2010). Within the tool's interface, we created two collections titled "interpreting" and "speech-to-speech translation".

Metadata obtained through PoP were then displayed in both collections, for editing and viewing, as shown in Figure 3. It allowed for the manual correction of inconsistencies resulting from the automatic processing of references by PoP or GS.

Figure 3 - Retrieved references information and metadata.

	AUTHORS	YEAR	TITLE	SOURCE	ADDE
☐	☆ Yang, Z; Shimizu, S; Zhou...	2023	The Kyoto Speech-to-Speech Translation System for IWSLT 2023	... Language Transl...	14/02
☐	☆ Sperber, M; Neubig, G; Ni...	2019	Attention-passing models for robust and data-efficient end-to-end speech tr...	Transactions of the ...	14/02
☐	☆ Reithinger, N; Maier, E; Al...	1995	Treatment of incomplete dialogues in a speech-to-speech translation system		14/02
☐	☆ Wang, C; Pino, J; Wu, A; ...	2020	Covost: A diverse multilingual speech-to-text translation corpus	arXiv preprint arXiv:...	14/02
☐	☆ Dong, Q; Ye, R; Wang, M; ...	2021	Listen, understand and translate: Triple supervision decouples end-to-end s...	Proceedings of the ...	14/02
☐	☆ Niehues, J; Salesky, E; Tu...	2021	Tutorial: End-to-End Speech Translation	Proceedings of the ...	14/02
☐	☆ Akbacak, M; Franco, H; Fr...	2009	Recent advances in SR'S IraqComm™ Iraqi Arabic-English speech-to-spe...	... Acoustics, Speec...	14/02
☐	☆ Tan, Z H; Lindberg, B; Ga...	2008	Handheld speech to speech translation system	Automatic Speech ...	14/02
☐	☆ Chung, Y A; Weng, W H; T...	2019	Towards unsupervised speech-to-text translation	... on Acoustics, Sp...	14/02
☐	☆ Peng, Y; Kulikov, I; Yang, ...	2024	MSLM-S2ST: A Multitask Speech Language Model for Textless Speech-to-S...	arXiv preprint arXiv ...	14/02
☐	☆ Pino, J; Puzon, L; Gu, J; ...	2019	Harnessing indirect training data for end-to-end automatic speech translati...	arXiv preprint arXiv ...	14/02
☐	☆ Bredmar, F	2017	Speech-to-speech translation using deep learning		14/02
☐	☆ Shankarappa, R T; Tiwari, S	2022	A faster approach for direct speech to speech translation	2022 IEEE Women I...	14/02
☐	☆ Ettelaie, E	2011	Concept classification with application to speech to speech translation		14/02

Source: Mendeley interface.

Many of the references retrieved in both corpora did not concern the areas of language interpreting and S2ST. Therefore, we decided to check each of the references and select those relevant to our study.

Due to the large number of references retrieved (264 papers in interpreting corpus and 953 papers in S2ST corpus), the work of checking and editing metadata for both corpora took 386 days to be completed. The information verified included: topic, authorship, year of publication, electronic address, and DOI (Digital Object Identifier, Paskin, 2010). We excluded all restricted access references from the collection.

Peer-reviewed papers were included if they directly addressed S2ST and interpreting, were published between 2014 and 2024, presented a clearly described methodology, and provided full-text access. Studies with a tangential focus, lack of methodological rigor, opinion pieces, or publications without peer review were excluded.

2.1.3. Downloading the academic publications

PDF files of each open access reference in both corpora were automatically downloaded using a Python script, available in Appendix I and developed by the author of this study with the assistance and debugging of ChatGPT model o3-mini-high³, specialized in logic and programming. The script searches and downloads open-access academic publications listed by GS based on the DOI identifiers retrieved through PoP and edited in Mendeley.

We conducted the search for academic publications using SerpAPI⁴, an API that enables automated searches on Google without dealing with blocks, captchas, or request limitations. It provides structured Google results as if they were data organized in JSON (Crockford, 2012). This study uses the free version of SerpAPI, which allows up to 100 searches per month and is for non-commercial use only.

The script available in Appendix I is the complete and refined version after 13 attempts. The main errors encountered during the development and execution of the code were related to syntax and DOI web addresses.

The automated process of downloading the academic publications took about three hours. At the end, we stored the downloaded academic publications on the computer. Due to time constraints, we renamed all files at once with the label *file_(number)*.

³ <https://openai.com/index/openai-o3-mini/>.

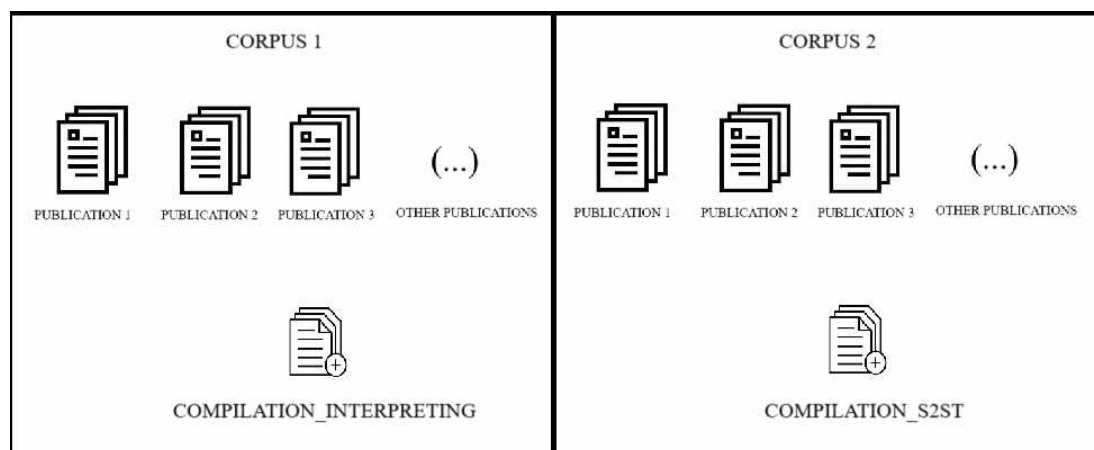
⁴ <https://serpapi.com/>.

2.2. Identifying the stages of development of S2ST systems and PIs' competence

After retrieving and downloading the academic publication files, ChatGPT-o1 retrieved information about the competence of PIs (corpus 1) and the stages of development of S2ST systems (corpus 2). The model we applied to such an extraction was o1⁵, specialized in advanced reflection.

At the time this study is being conducted, the maximum number of files that could be attached to ChatGPT-o1 is 10. Therefore, we compiled the PDF files retrieved in the previous steps into two files, as shown in Figure 4. We used Adobe Acrobat Pro⁶ for this compilation.

Figure 4 - Corpora compilation and organization.



Source: The author.

We utilized ChatGPT-o1 Plus, a paid subscription AI service, due to its enhanced messaging and file upload capabilities. The ChatGPT-o1 application version we utilized was 1.2025.41.

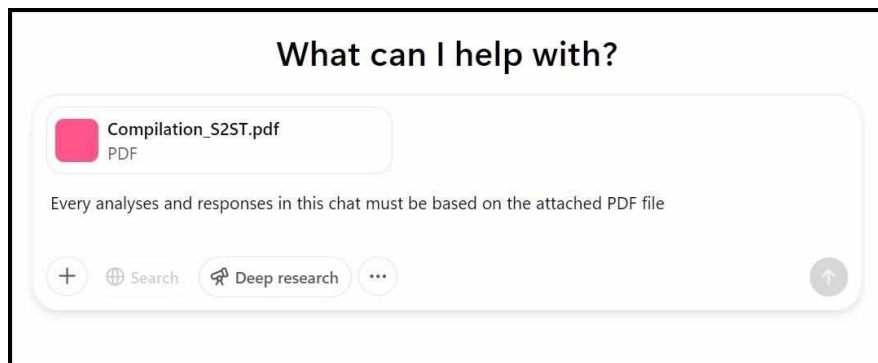
⁵ <https://openai.com/o1/>.

⁶ <https://www.adobe.com/br/acrobat.html>.

Plus subscription facilitates data analysis and access to extensive research and diverse reasoning models. Additionally, for the purposes of this study, it enabled the creation and utilization of “projects”⁷, a feature that allows for the allocation of multiple analyses.

We attached the files containing all retrieved publications for both corpora to two different chatboxes. As an example, Figure 5 shows the Compilation_S2ST file (corpus 2) attached to the chatbox in the ChatGPT interface.

Figure 5 - Retrieved publications attached, advanced reasoning model.



Source: ChatGPT interface.

We selected the advanced reasoning o1 model. Then we gave the prompt “All of the analyses and responses in this chat must be based on the attached PDF file”.

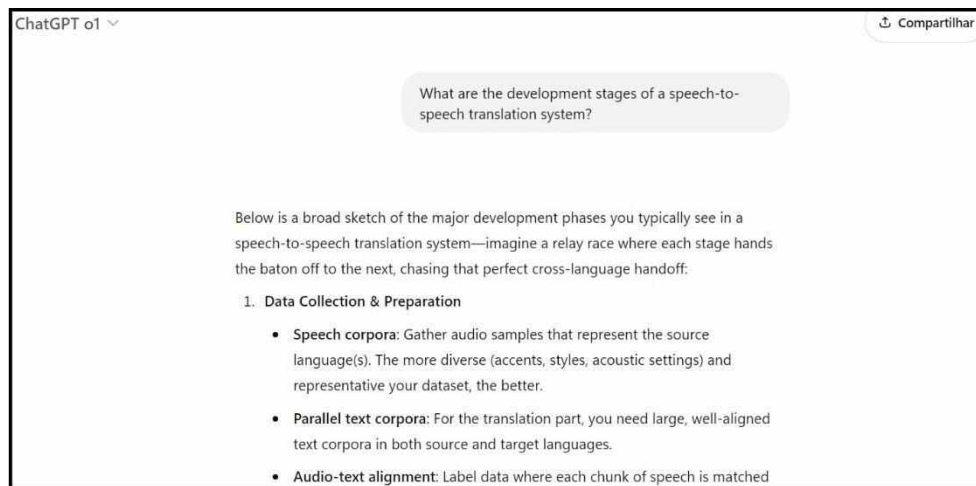
After giving this prompt, the AI tool responded: “All right! I’m all set to dive into the PDF and keep our conversation laser-focused on its content. Feel free to ask away about anything that piques your curiosity in the attached document — I’m at your service!” (ChatGPT).

Next, the prompts we gave to each chatbox: (1) “Based on the academic publications attached to this chat, what knowledge and skills build the competence of a professional interpreter?”; and (2) “According to the academic publications attached to this chat, what are the stages of development of S2ST systems?”.

After giving the prompts in two separate chatboxes, ChatGPT-o1 produced two reports, one related to the academic publications on Interpreting and the other related to the development stages of S2ST systems. The reports contain text and excerpts of text extracted from academic publications, in narrative style and in topics, as show in Figure 6.

⁷ <https://help.openai.com/en/articles/10169521-using-projects-in-chatgpt>.

Figure 6 - Report generated by ChatGPT.



Source: ChatGPT interface.

Given that, we gave the prompt “include the references for each excerpt of extracted text”. We did this in both chatboxes and the reports were updated with the references to the corpus files.

Finally, we attached the two reports to a single ChatGPT-o1 chatbox and submitted the prompt: “Based on the two reports attached to this chat, what are the stages of development of S2ST systems to which PIs can contribute to?”. The findings are discussed in chapters 6, 7 and 8.

PART I

3. Related work

The main focus of this study is to understand the role of PIs within S2ST in addition to data supplying and output evaluation. We identify the knowledge and skills that comprise the competence of a professional interpreter according to academic literature. We also identify the stages of development of S2ST systems.

Our study employs technological tools for the retrieval of academic publications focused on Interpreting and S2ST. Data acquired from these publications are analysed by AI through advanced reflection models that analyze semantically and contextually based on the prompts given in the chatbox.

We are not aware of any previous work addressing this subject by using AI. However, this work is related to and stems from ideas presented by Franz Pöchhacker in his works on the intersections of Translation, Interpreting and technology.

3.1. Interpreting and technology

The precise origin of human oral language remains one of the most significant scientific enigmas. Due to its ephemeral and intangible nature (Smeets, 2004), researchers from various disciplines rely on indirect evidence and diverse theories to understand its emergence (Dessalles, 2007).

Current estimates vary considerably: from 2 million years, with *Homo habilis*, to approximately 100,000 years ago, with *Homo sapiens* (Holloway, 1983). These estimations were obtained through indirect evidence such as the analysis of the anatomy of fossils, specifically the larynx and brain regions.

There are various theories about the origins of language, ranging from those that support a gradual evolution to those that propose an abrupt emergence (Yule, 2022). However, many scholars agree that oral language is a technology that has been utilized and manipulated by the human species since time immemorial.

Oral language, akin to any technological advancement, is a tool that facilitates the construction of connections, the exchange of ideas, and the shaping of reality. Similar to all other forms of technology, oral language evolves over time, modernizing and adapting to the needs and social contexts (Farrell, 1985).

Given this, PIs' use of technology is not an entirely new phenomenon. They have utilized such technology from the outset of their profession. The use of electronic technologies by PIs, however, only began in the 1920s, with the emergence of electroacoustic sound transmission systems that enabled the development of simultaneous interpreting (Baigorri-Jalón, 1999).

Braun (2019) identifies four distinct types of technology currently accessible to the professional interpreter:

[...] technologies used to deliver interpreting services and enhance their reach, leading to technology-mediated or distance interpreting; technologies that can be applied to support or enhance the professional interpreter's preparation, performance and workflow, leading to technology-supported interpreting; and technologies that are designed to replace professional interpreters, leading to technology-generated or machine interpreting. A further category that could be added is that of technology-enabled hybrid modalities such as respeaking [...]. (Braun, 2019, p. 271)

According to Braun (2019, p. 271), these technologies will have a more profound and groundbreaking impact on PIs, their clients, and the entire interpreting ecosystem than any technological advancements over the past few decades. Fantinuoli (2018) calls this phenomenon "the upcoming technological turn".

Pöchhacker has been monitoring the growing role of technology in interpreting. In his academic works he addresses the topic of technological advances typically as part of the broader discussion around interpreting research, practice, and training (Pöchhacker, 2013).

S2ST technology is rapidly improving, but researchers tend to see it as a complementary tool rather than a full-scale replacement for PIs. Pöchhacker (2024), for example, remarks that

It seems unlikely that it will ever be possible to mechanize meaning-making as an embodied (bio)semiotic process the way, say, traffic control by police was replaced by traffic-light systems. Until this happens, speech translation by machine can nevertheless be a helpful instrument approximating the functionality of interpreting, which is to enable communicative interaction and access to information (p. 18).

Pöchhacker (2006) raises the question of whether S2ST can be entrusted with high-stakes settings — like legal or healthcare contexts — where the ethical and emotional dimensions are enormous. Later, consistent with his broader perspective on interpreting

quality, Pöchhacker (2008) emphasizes that interpreting is not just about words but also cultural context, interpersonal interaction, and empathy — traits that are tough to include in algorithms.

On a didactic level, Pöchhacker (2013) suggests that emerging technologies might affect how future PIs are trained. Students may need to be prepared to work with and alongside these tools, evaluating their outputs, troubleshooting issues, or even “post-editing” spoken machine translations (similar to post-editing in written machine translation).

As a whole, Pöchhacker’s perspective on S2ST is balanced: he does not dismiss the rise of machine interpreting⁸, yet he remains constructively skeptical. He questions whether algorithms, with their growing but still bounded linguistic competencies, will ever be able of fully capturing the interpretive finesse and emotional intelligence of PIs.

Another important scholar investigating this topic, Horváth (2021) examines the evolution of S2ST, contrasting it with human interpreting. She emphasizes that, while AI-based S2ST systems have advanced, they differ fundamentally from the intricate processes PIs employ. In her words,

Automated speech translation has an undeniable role and place in today’s world, steeped in technology and AI. However, it needs to be underlined that it is completely different from the complex interpreting service professional interpreters provide and the circumstances and contexts in which its use can be advised is intrinsically different from that of human interpreting. (p. 174)

Horváth (2021) advocates for a clear distinction between S2ST and human interpreting, suggesting that the term “interpreting” be reserved for human activities, while “automated speech translation” describes computer-mediated language transfer as the one performed by S2ST systems.

Beyond the technical aspects, Horváth (2021) explores the ethical considerations arising from AI’s integration into interpreting. In the paper *AI in interpreting: Ethical considerations* (2022), she discusses how AI and machine learning technologies impact data protection and confidentiality within the language mediation market.

In a webinar titled *With a little help from my friends: AI for professional interpreter training and support*, Horváth (2021) presents various AI tools available to PIs, focusing on terminology management systems and their potential to enhance both preparation and performance during assignments. Horváth’s (2021) insights contribute to an understanding of

⁸ Pöchhacker and some other scholars refer to S2ST as “machine interpreting” (Pöchhacker, 2024).

how AI can complement PIs, enhancing efficiency while maintaining the quality and ethics of the profession.

Interpreting used to depend entirely on the physical presence of the professional interpreter (Donovan, 2017). With technological advances, however, new practices have transcended the boundaries of the traditional profession.

Now many new technological innovations have been used, moving the professional interpreter away from the interpreting site (Saina, 2021). In a way that it is now possible for interpreting to be done in contexts which would previously have been unthinkable because of distance and lack of resources (Corpas Pastor, 2021).

Despite these opportunities, the increased cognitive load and stress faced by PIs when working remotely are significant issues. Moser-Mercer (2005) remarks that

The lack of virtual presence has emerged as one of the major factors determining poorer performance in remote as opposed to live simultaneous interpreting. This deterioration of quality appears to be based in early onset of fatigue, which in turn seems to be a consequence of allocating additional cognitive resources to comprehension processes during simultaneous interpreting and therefore depriving other parts of the process, notably production, of the resources necessary to maintain a high level of performance during normal turn time. (p. 226)

Technological tools have also impacted professional interpreter training, offering dynamic and interactive learning environments. Virtual reality (VR) and simulation software can now recreate authentic interpreting scenarios, exposing trainees to realistic conditions in safe and controlled virtual spaces (Braun et al., 2013).

Such immersive technologies provide students with valuable practice, instant feedback, and the opportunity to master coping strategies for real-world interpreting challenges. Consequently, technology not only enhances PIs' professional preparedness but also allows continuous skill refinement throughout their careers (Zhang; Liu, 2024).

While technology expands the accessibility of interpreting services, it also raises critical ethical and professional questions. Remote interpreting can create confidentiality concerns, since sensitive information traverses digital channels prone to cybersecurity risks (Kushalnagar et al., 2019).

Technology has undoubtedly improved accessibility and inclusion within interpreting, empowering communities previously underserved due to geographical isolation or economic

limitations. In this sense, Almahasees et al. (2024) investigates the perceptions of remote PIs regarding the shift from on-site to online interpreting modes during COVID-19 pandemic.

Almahasees et al. (2024) highlight how platforms like Zoom, telephone services, and Kudo facilitated continued interpreting services, particularly for healthcare providers and international organizations. They also address challenges such as technological limitations and the lack of non-verbal communication, offering recommendations for integrating remote interpreting into crisis response strategies.

3.2. PIs as the reference standard

Researchers who develop S2ST systems often measure their progress by comparing outputs to those of PIs. According to Pöchhacker,

(S2ST) evaluation benchmarked against human interpreting performance constitutes a significant link between the computational translation research community and interpreting scholars. Nevertheless, the present state of the art is characterized by a lack of interaction and exchange, and possibly by different conceptualizations much more than discrepancies in terminology. This leads back to the question posed in my title, addressed below (2024, p. 10, emphasis mine).

This shared focus on human benchmarks is the main bridge connecting the computational translation community and interpreting scholars since both groups care about how closely the machine's performance matches human's performance. Despite this common goal, Pöchhacker (2024) observes that genuine collaboration between the two fields (Interpreting and S2ST) is rare.

PIs' ability to understand context and cultural aspects sets a high standard for automated systems to emulate. Studies have highlighted that PIs excel in detecting contextual clues, a skill where S2ST translation systems frequently falter (Hudelson & Chappuis, 2024).

Despite advances in AI, MT systems — a core technology in S2ST systems — struggle with figurative meaning and cultural references. Alawi and Abdulhaq (2017) examine the limitations of MT systems in accurately translating idiomatic expressions and culturally specific content.

Their study highlights that, while MT can provide general translations, it often fails to convey subtle meanings. PIs, however, tend to be able to cope better with these challenges (McDonald, 1981).

Another problem is that in S2ST systems, the latency of MT output interferes with the flow and rhythm of conversation. In contrast, PIs provide more or less instantaneous renderings and smoother interactions (Horváth, 2021).

Machines can process vast amounts of data, but they lack human's ability to adapt to unexpected shifts in conversation dynamics. Such an adaptability is essential for maintaining coherence and relevance during dialogues (Wester et al., 2010).

PIs possess the ability to manage ambiguity and infer meaning from incomplete information. Machine systems, however, require explicit input to function properly, which limits their effectiveness in spontaneous conversations (Prasad et al., 2012).

Another important element is empathy. PIs can feel and respond to emotions, customizing their deliveries to suit the emotional context, a capability currently beyond any machine (Korpal & Jasielska, 2019).

In diplomacy and business, the precision and reliability of PIs are indispensable. Their competence ensures that meaning is accurately conveyed, preventing potential misunderstandings (Kadric, 2021).

PIs are also able to clarify and receive feedback on their messages from others, offering a more collaborative approach to the quality of communication (Wadensjo, 2014). Machine systems, lacking this interactive capability, may maintain indefinitely errors without correction (Sperber, 2020).

Next section deals with how ChatGPT and similar AI tools can support academic research by performing semantic analysis, summarizing literature, and organizing large datasets. While these tools enhance productivity and trend detection, scholars emphasize the need for human oversight due to risks of inaccuracy and fabricated content.

3.3. Artificial intelligence-based semantic and contextual analysis

This study analysed data collected from academic publications using ChatGPT, an advanced generative artificial intelligence⁹ system developed by OpenAI¹⁰. It employs deep learning techniques derived from transformer-based Neural Networks (NNs) to produce coherent and contextually relevant text outputs (Hashana et al., 2023).

⁹ Within this study, the term “artificial intelligence,” alongside its acronym “AI,” is employed as a synonym for “conversational assistants” and “chatbox”.

¹⁰ <https://openai.com/>.

A simple search on GS reveals that from 2014 to 2024 around 4,600 academic papers on S2ST were published. This quick and superficial search, however, only includes works indexed in English.

In such a scenario of increasing complexity, S2ST research necessitates advanced tools for semantic analysis. Large language models like ChatGPT offer potential solutions for processing and understanding academic publications in this field (Deng & Lin, 2023).

Semantic analysis involves extracting meaning from text to identify relationships between concepts. ChatGPT, trained on extensive datasets, can recognize patterns and interpret contextual relationships within academic discourse (Titus, 2024).

Researchers often struggle with the vast volume of academic publications. AI can assist by summarizing key findings and categorizing themes across multiple sources (Turobov et al., 2024).

Traditional literature reviews require extensive manual effort to synthesize data. By leveraging ChatGPT, researchers can speed up this process and focus on critical analysis rather than mere data collection.

Haman & Školník (2024) wrote an open letter about the ethical and practical implications of using ChatGPT-o1 for literature review. The researchers report having used ChatGPT in 2023 and, according to them,

ChatGPT was given the prompt [...]: *“List 10 seminal academic articles in the field of medicine and provide the DOI.”* We used this prompt five times, resulting in five unique chats. ChatGPT provided us with 10 publications each time, for 48 unique publications from the five attempts. We found that only 8 of the 50 provided DOIs exist and are listed for the correct publication. At the same time, we searched for the title of the paper itself with the authors and the correct academic publication (even though the DOI may have been wrong), and only in 17 out of 50 cases did the article in question exist in the databases (Google Scholar, PubMed, Semantic Scholar). In 66% of the cases, ChatGPT gave us a non-existent paper. Considering that the prompt was general to the whole field of medicine and not a prompt on specific research questions, these are very poor results. ChatGPT might have given much worse results if we had wanted to focus on a specific research question. (p. 1244)

Khalifa and Albadawy (2024), however, present a systematic review of 24 studies from major databases since 2019, highlighting AI's impact on academic writing and research. According to their study, AI enhances academic writing in six areas: idea generation, content structuring, literature synthesis, data management, editing, and ethical compliance.

Moreover, according to their findings, ChatGPT demonstrates significant potential in academic writing, though challenges in academic integrity and AI-human balance persist. That is, human oversight remains essential to ensure accuracy and interpretability (Mondal & Mondal, 2024).

ChatGPT's natural language processing (NLP) capabilities enable it to detect trends in terminology and emerging research directions. This allows scholars to identify gaps in knowledge and potential areas for further study.

Hsu (2024), for example, conducts an exploratory study with 60 participants to examine whether the utilization of AI enhances students' understanding of medical terminology. The research findings indicate that incorporating AI as learning assistants can significantly enhance the learning process. This was evidenced by a substantial improvement in post-test scores among participants in the experimental group (who used AI learning assistants) compared to those in the control group.

Categorizing things is part of a researcher's job. Concept clustering is an AI feature that organizes the subjects of a publication or set of publications into categories. It is useful because, instead of searching through everything manually, researchers can quickly find related topics and see patterns in information (Singh & Singh, 2024).

Another AI feature that can make researchers' lives easier is the ability to analyse abstract structures in academic papers, allowing the identification of research methodologies and results (Pal, 2023). This can help scholars compare experimental designs and evaluation metrics across studies.

Kardas et al. (2020) propose AXCell, a tool that automatically reads academic papers and pulls out the most important results. It reads the papers, finds tables with useful results, figures out what each number means and organizes results so researchers can compare results easily.

AI can also facilitate keyword extraction, ensuring that critical concepts are properly indexed (Chan et al., 2023). This benefits researchers searching for relevant literature in extensive databases.

And this type of extraction is also used in other contexts. Lee and Kim (2008), for example, present a methodology to automatically find the most important keywords from huge amounts of online news articles.

Since thousands of news stories come out every day, it is impossible for people to manually read and summarize them. So, Lee and Kim (2008) created a system that looks

through all the news, picks out the most meaningful words, and helps group related stories together by topic.

One challenge of using ChatGPT is its occasional tendency to generate plausible but incorrect interpretations. Cross-referencing with original sources is necessary to validate findings.

Kim et al. (2025) examine how mistakes made by ChatGPT can affect whether people trust its travel advice and want to follow it. Researchers run six experiments to see what happens when ChatGPT gives wrong information — especially how obvious the mistake is and how closely it is related to the travel topic.

The study found that when there were no mistakes, people felt the advice was accurate and trustworthy, and they were more likely to want to visit the place ChatGPT recommended. But when ChatGPT made errors — especially ones that were easy to notice or clearly about travel — people were less likely to follow the advice.

If the mistakes were about unrelated topics or if users were really focused on their main goal, the negative effect was not as strong. That is, people are less likely to trust or act on AI travel tips if the AI messes up in obvious or relevant ways.

The scalability of ChatGPT makes it a powerful tool for analysing large datasets of publications (Mechkaroska et al., 2024). This is particularly useful for meta-analyses and reviews.

Geng and Trotta (2024) evaluate one million paper abstracts submitted to the website arXiv between 2018 and early 2024. Their objective is to find out how much ChatGPT and similar AI tools are influencing how scientists write.

To do this, Geng and Trotta (2024) use statistics to track how the choice of words changed over time, and compared real abstracts to ones that had been rewritten by ChatGPT. Analysis shows that a growing number of papers — especially in computer science — are likely written or edited using ChatGPT-like tools.

They estimate that in computer science about one-third of recent abstracts now follow a writing style typical of ChatGPT. Geng and Trotta (2024) end by discussing the pros and cons of this trend: while these tools can improve clarity and accessibility, they may also flatten diversity of expression and raise concerns about originality.

The combination of ChatGPT with human expertise results in a hybrid approach to semantic analysis (Szandała, 2025). This synergy can improve the depth and accuracy of literature reviews.

Yan et al. (2024) experiment with combining ChatGPT with human expertise to help analyse the meaning of large amounts of written information. Instead of doing all the work manually, which takes a long time, they let ChatGPT help by suggesting patterns, grouping similar ideas, or pointing out repeated themes in the text.

Humans are still needed to guide ChatGPT, check its suggestions, and make sure the analysis makes sense in context. In this sense, ChatGPT is like an assistant which is good at organizing messy notes quickly, but it still needs a knowledgeable person to review and decide what is truly important or meaningful.

As an assistive tool, ChatGPT can enhance academic productivity by reducing the cognitive load associated with manual data synthesis. However, researchers must remain vigilant in ensuring scholarly rigor (Khalifa & Albadawy, 2024).

The refinement of ChatGPT's capabilities through fine-tuning on specialized datasets can improve its accuracy in domain-specific analysis. Tailored training data may mitigate generalization errors (Nishat & Mustafa, 2024).

4. Who is the PI?

The central theme of our study is PIs. Our goal is to understand the role of PIs within S2ST in addition to data supplying and output evaluation.

Descriptions of who the professional interpreter is have been written by several scholars of Interpreting Studies. In this work, following the academic trend presented in Chapter 3, we gave ChatGPT-o1 the prompt “who is the professional interpreter?” with interpreting corpus attached to the chatbox.

In our search, we used the “deep search” feature, asked it to focus on the area of Interpreting Studies and attached the interpreting corpus composed by 264 academic papers.

As soon as we gave the prompt, ChatGPT-o1 requested more information, as shown in Table 1. As our objective in this chapter is to present a brief overview of PIs’ professional identity, we answered to the request with the prompt “we would like to know about the historical evolution of the professional interpreter’s role and about interpreters’ social, ethical, and professional identity”.

Table 1 - ChatGPT request for more information.

To best answer your question on the professional interpreters and their role, could you specify which dimension you’re most interested in? For example:

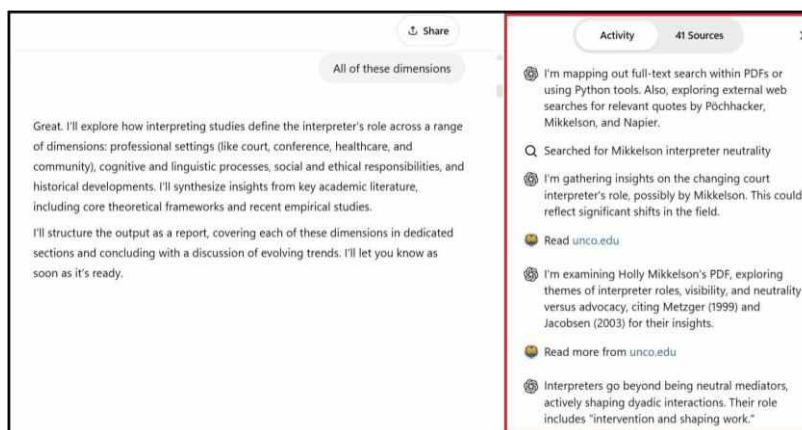
- The professional interpreter’s role in specific settings (e.g. court, healthcare, conference, community)
- The cognitive or linguistic aspects of the professional interpreter’s work
- The professional interpreter’s social/professional identity or ethics
- Historical evolution of the professional interpreter’s role

Let me know your desired angle so I can tailor the summary accordingly from the Interpreting Studies perspective.

Source: ChatGPT interface.

Then, the search began to be carried out. While ChatGPT-o1 was “reasoning”, on the right side of the screen it was possible to visualize the paths traced by the AI search tool, as illustrated in Figure 7.

Figure 7 - ChatGPT research activity.



Source: ChatGPT interface.

The search took 18 minutes to complete and consulted 41 sources of which only 10 were reported. The result presented is a report with 1,352 words and whose structure is shown in Table 2.

Table 2 - Structure of the report generated by ChatGPT.

Title: The historical evolution of the professional interpreter's role and on the social, ethical, and professional identity of interpreters
1. Historical Evolution of the Professional Interpreter's Role - Ancient and pre-modern eras - The birth of conference interpreting - Professionalization and training - Interpreting Studies as a field
2. Social, Ethical, and Professional Identity of Professional Interpreters - Neutrality vs. advocacy - Ethical frameworks - Professional identity - Situational and social context

Source: ChatGPT-o1 interface.

The accuracy of the report generated by ChatGPT-o1 was manually verified against the corpus of academic publications on Interpreting. In passages where the tool did not provide a

reference, we searched the corpus using keywords extracted from the passage itself and inserted the corresponding reference.

Due to the space limitations of this work, we will not discuss the organization of the report items and references recovered. The literature review that follows is based on the report generated by ChatGPT-o1, whose full version is found in Appendix II.

4.1. Historical evolution of the professional interpreter's role

According to ChatGPT-o1's findings, the role of PIs has changed dramatically from ancient times to the present, shifting from untrained individuals to highly trained professionals (Delisle & Woodsworth, 2012). This evolution is tied to historical needs (e.g. trade, diplomacy, colonization, international relations) and to the emergence of interpreting as a recognized profession and academic discipline (Biagini et al., 2017).

Telling the story of interpreting profession is also, in a way, telling the story of interpreting itself. On telling this story, Baigorri-Jalón (2015) asks: "Is it then possible to write a history of interpreting and, if so, what for and how?" (p.11).

According to the researcher, this undertaking is only possible "if we look in the primary sources for the function of interpreting rather than the current concept of the profession" (p.11).

4.1.1. Ancient and pre-modern eras

Interpreting is often cited as "the world's second oldest profession" (Zielnik-Kołodzińska, 2021). Wherever multilingual contact occurred, there was a need for someone to facilitate communication. In ancient civilizations, PIs already held important if precarious positions.

For example, Ancient Egyptian records from the 3rd millennium BCE mention officials designated as interlingual mediators in the Pharaoh's court (Piller, 2021). Egyptian princes at Elephantine had "overseers of dragomans" – indicating that PIs existed as a recognized group (dragoman being a historical term for professional interpreter) and even had hierarchy and oversight (Kurz, 1985).

The Bible offers an early reference in Genesis 42: Joseph converses with his Hebrew brothers through a professional interpreter while he is an official in Egypt (Piller, 2021). This underscores that PIs were used in governance and trade in antiquity.

In the Greek and Roman eras, PIs (often slaves or bilingual merchants) were vital in diplomacy and war (Pepiler, 1915). According to ChatGPT, “however, Latin proverb ‘*interpres nullus amicus*’, which means ‘no professional interpreter is a friend,’ demonstrates that those professionals were sometimes subject to mistrust and suspicion of their loyalty to authorities”.

In our inspections, however, we did not find any source available online directly documenting such Latin expression. The closest we got to it is Ovid’s *Epistulae ex Ponto* (Letters from the Black Sea), Book IV¹¹, where the Roman poet laments that a malicious professional interpreter has incited anger against him by misrepresenting his words:

“At malus interpres populi mihi concitat iram inque novum crimen
carmina nostra vocat” (Ovid, undated).¹²

Throughout the Middle Ages, PIs facilitated contact during travel, missionary work, and exploration. Notably, during the Age of Discovery and colonization, individuals like La Malinche, an American native woman who interpreted for Cortés in the Aztec Empire (Valdeón, 2013) or Sacagawea, who interpreted for Lewis and Clark in Canada (Karttunen, 1994), were essential as cultural and linguistic mediators.

These PIs often learned languages in informal ways and served under extreme power imbalances. Sometimes they were revered for their utility, other times reviled as traitors. Indeed “traitor” and “translator” share an etymology in some languages (Mairs, 2011).

4.1.2. *The birth of conference interpreting*

The early 20th century saw a turning point with the rise of international diplomacy requiring formal multilingual communication (Nick, 2001). Until World War I (WWI), diplomatic meetings were often conducted in a single lingua franca (French), minimizing need for PIs (Wilss, 1999).

After WWI, however, the 1919 Paris Peace Conference introduced English alongside French as a diplomatic language (Baigorri-Jalón, 2014). This led to the use of consecutive PIs for the first time in high-level international negotiations. Figures like Paul Mantoux became

¹¹ <https://www.thelatinlibrary.com/ovid/ovid.ponto4.shtml>? Retrieved 24 March 2025.

¹² This translates to: “But a malicious professional interpreter of the people excites anger against me and calls my poems a new crime.”

famous for interpreting speeches by statesmen sentence by sentence to the other language (Pöchhacker, 2022).

The League of Nations (1920s) also employed PIs, mostly consecutive or whispered interpretation (Morris, 1997). These PIs were pioneers, developing techniques for note-taking and public speaking in another language without a script.

Still, interpreting was not yet institutionalized as a full-fledged profession. There were only a handful of practitioners, often diplomats themselves or individuals with unique bilingual upbringing (Baigorri-Jalón, 2015).

A critical moment came with the Nuremberg Trials (1945–46) after WWII (Gaiba, 1998). Given the scale and complexity of the events, traditional consecutive interpretation was deemed impractical, as it would have significantly prolonged the prosecutions (Baker & Diriker, 2019).

In order to run the trials efficiently in four languages, the organizers set up a system that used headsets, microphones, and IBM equipment, which allowed PIs to interpret speech into other languages instantly (Guise, 2020). This was the first major use of simultaneous interpreting in history.

The success at Nuremberg led the newly formed United Nations (UN, est. 1945) to adopt simultaneous interpretation for its assemblies and conferences (Baigorri-Jalón et al., 2021). In the following decades, demand for skilled conference PIs grew with the Cold War (Fernández-Sánchez, 2019).

4.1.3. Professionalization and training

With conference interpreting emerging as a distinct occupation in the 1950s, steps were taken to formalize the profession (Setton & Dawrant, 2016). The International Association of Conference Professional Interpreters (IACI or AIIC, where the acronym AIIC comes from the French name: Association Internationale des Interprètes de Conférence) was founded in 1953, setting forth a code of ethics, working conditions, and standards for recruitment (Longley, 1984).

AIIC helped solidify conference interpreting as a respected profession with global reach (Keiser, 1999). Around the same time, specialized professional interpreter training schools were established: the Ecole Supérieure d'Interprètes et de Traducteurs (ESIT) in Paris, the interpreting program in Geneva, and later schools in Heidelberg, Monterey, etc. (Pöchhacker, 2012).

These programs developed curricula focusing on consecutive and simultaneous techniques, domain knowledge, and professional ethics. Graduates of such programs became the new generation of PIs staffing of the UN and European institutions.

Throughout the mid to late 20th century, the professional interpreter's role was largely conceived as that of a highly skilled language specialist who should be as discreet as possible. Seleskovitch's Interpretive Theory of Translation (ITT) in the 1960s–70s (Tryuk, 2023) emphasized the professional interpreter's understanding of meaning but still viewed the professional interpreter as a non-participating narrator of someone else's speech.

Meanwhile, court interpreting and community interpreting lagged behind in professional development. Court interpreting in the US, for instance, was not systematically addressed until the 1970s, when the needs of non-English-speaking defendants led to the Federal Court Professional Interpreter Act (1978) and the creation of certification programs (Barcelo Jr., 1979).

Before that, court PIs were often untrained volunteers or off-duty bilingual staff. Similarly, community interpreting (in hospitals, etc.) was often done by whoever was available (family members, bilingual staff) well into the late 20th century (Mikkelsen, 1999).

During the 1990s, a growing awareness of the rights of immigrants and the importance of equal access to services led many countries to invest in substantial changes to their language support systems (EBSCO, 2024). This included significant investments in training programs specifically designed for community PIs, as well as the development and implementation of professional codes of practice to guide their work (Niska, 2008).

4.1.4. Interpreting Studies as a field

The scholarly understanding of PIs' role also evolved. Early writings (1950s–70s) on interpreting were mostly practical manuals or memoirs by conference PIs, often reinforcing the ideal of the professional interpreter as a neutral messenger (Pöchhacker, 2022).

In the 1990s and early 2000s, however, researchers like Cecilia Wadensjö (1992), Ebru Diriker (1998), and Claudia Angelelli (2001) started applying sociolinguistic and discourse-analytic approaches to professional interpreter-mediated encounters. This gave rise to a more diverse picture of PIs as co-constructors of communication, not invisible conduits (Valero-Garcés, 2007).

Historical perspective reveals a shift: where once the “conduit model” was unassailable dogma, by the 21st century it became clear that PIs have some agency and their role is negotiated with other participants. For example, historical transcripts of interpreted events when analysed show PIs sometimes intervened or influenced outcomes (Baigorri-Jalon, 2010).

In terms of status, PIs have gone from being “necessary evils” to being seen as essential professionals in international and multicultural settings (Nachman, 2020). A telling sign of this evolution is the institutionalization.

Today many countries legally require qualified PIs in courts and public services. One clear example of this is the Directive 2010/64 (European Union, 2010) on the right to interpretation in criminal proceedings, something that would have been rare a century ago.

All these advances culminated in the creation and establishment of Interpreting Studies (IS). An academic discipline that emerged formally only in the 20th century, driven by the professionalization of interpreting.

The IS have grown into a comprehensive discipline of research drawing from linguistics, sociology, psychology, communication studies, and translation theory. According to Pöchhacker (2022), the IS must be understood as a map of interrelated domains and dimensions, including the professional interpreter’s role, institutional contexts, cognitive processes, and ethical concerns.

Initially, interpreting was considered a subset of Translation Studies (TS); however, the unique challenges and skills associated with real-time language transfer necessitated specialized research and theoretical frameworks (Gile, 2006). This led to the recognition of Interpreting Studies as a separate discipline, addressing diverse modes such as simultaneous, consecutive, and community interpreting (Gile, 2009).

Several universities have established dedicated faculties and research centers for Translation and Interpreting Studies (TIS), reflecting the field’s academic maturity. Notable among these are Faculty of Translation and Interpreting at the University of Geneva (Ramos, 2015) and Faculty of Translation and Interpreting of the University of Granada (Baxter, 2019).

Contemporary research in IS explores areas such as the impact of technology on interpreting practices (Chan, 2024), the role of PIs in multicultural settings (Yi, 2024), and the cognitive mechanisms underlying interpreting performance (Moratto & Yang, 2024).

4.2. Social, ethical, and professional identity of interpreters

According to the findings of ChatGPT, “professional interpreters do not work in a vacuum; they are social actors embedded in communicative events that often involve disparities in power, culture, and expectations”. Although ChatGPT does not inform the exact source of this maxim, expressions such as “social actors,” “communicative events,” and “disparities in power” give us a clue.

This maxim is supported by a convergence of sociological, pragmatic, and critical theories, which together defend a non-instrumental view of interpreting. It rejects the myth of absolute neutrality and positions the professional interpreter as a situated, culturally informed, and ethically involved mediator (Baker, 2018; Hale, 2014; Mason & Ren, 2014).

Over time, a rich discussion has evolved in Interpreting Studies around the social role of the professional interpreter (Biagini et al., 2017), the ethical principles guiding their decisions (Kalina, 2015), and how interpreters construct their professional identity (Rudvin, 2015). Key debates revolve around whether a professional interpreter should be a neutral, “invisible” facilitator or an active co-participant, and how PIs manage their allegiance to each party and to the message itself (Bronkhorst, 2011; Martínez-Gómez, 2015).

4.2.1. Neutrality vs. advocacy

Traditionally, the prevailing view – especially in conference and legal interpreting – was that PIs should be neutral bridges, with no stake or agency in the interaction (Metzger, 1996). This “conduit model” likens the professional interpreter to a telephone wire, transmitting information without distortion or influence (Chandler, 1994).

Codes of ethics in many settings explicitly call for impartiality and instruct PIs not to offer personal opinions or counseling (Kalina, 2015). For example, court PIs are “forbidden to give advice or provide explanations” beyond the direct interpretation (Mikkelsen, 2008, p. 73).

The professional interpreter’s voice is ideally transparent, and their presence minimal. However, scholarly research has challenged the notion of the completely “invisible” professional interpreter (Wadensjö, 1992).

In various real-world studies, it became evident that PIs inevitably shape the communicative process. They do it through their choices, and all parties, including PIs themselves – recognize this to some degree (Wadensjö 1998; Metzger 1999; Angelelli 2004).

Angelelli’s (2011) survey of conference, court, medical, and over-the-telephone interpreters found that PIs did not consider their role to be invisible in any setting. According to the findings PIs felt they “played a role in building trust, facilitating mutual respect,

communicating affect as well as message, explaining cultural gaps, controlling the communication flow and aligning with one of the parties” (p. 26). This finding highlights that PIs often see their responsibilities as going beyond rote interpreting – they are managing relationships and understanding between people.

The idea of alignment is particularly contentious: does a professional interpreter ever “ally” with one side? (Davidson, 2001). While codes insist on neutrality, in practice PIs may empathize or identify more with one party, often the weaker or non-dominant party (Bolden, 2000), such as a patient or asylum seeker, which can subtly influence their renditions (Halley, 2020).

For instance, a professional interpreter might soften a harsh reprimand from a doctor when conveying it to a patient, out of compassion, acting in a way as the patient’s advocate. Such interventions raise ethical questions (NCIHC, 2021).

Some scholars (like Valero-Garcés, 2003) argue that strict neutrality can actually be unethical if it leads to misunderstanding or harm; they suggest advocacy in the sense of making sure the marginalized voice is fully heard and understood. Others caution that advocacy undermines the professional interpreter’s credibility and could skew the interaction (NCIHC, 2003).

The consensus in recent literature is that PIs occupy a “third space” (Parenzo & Schuster, 2020). They are neither completely neutral nor advocates in the way a participant would be, but they have a mediating role that involves active decision-making. As Jacobsen (2004) observed in legal settings, PIs often focus on faithfully conveying speakers’ intended meaning, even if it requires additions or adjustments, which contradicts the pure conduit ideal.

4.2.2. Ethical frameworks

The professional ethics of interpreting typically cover principles like accuracy, confidentiality, impartiality, and professional demeanor (Phelan, 2019). These are enshrined in codes by bodies such as AIIC (for conference PIs) (Mikkelsen, 2000), RID (for sign language PIs) (Sheneman, 2016), and NAJIT (for judiciary PIs) (Baixauli-Olmos, 2020).

For example, confidentiality means a professional interpreter should not disclose information from an assignment (Kalina, 2015). Impartiality means they should not take sides or alter messages to favor any party (Ciordia, 2017).

Accuracy means everything said should be conveyed to the best of the professional interpreter’s ability, even if it is embarrassing or damaging to the speaker (Gieshoff & Albl-

Mikasa, 2024). However, the implementation of these principles can be complex and require subtle distinctions.

Ethical dilemmas arise, for instance, when a literal accurate interpretation would cause confusion: is it more ethical to clarify (thus deviating from verbatim accuracy) or to stick to the words (and allow a possible misunderstanding)? Different codes give different guidance.

The Australian AUSIT¹³ code allows PIs to explain a term if needed, whereas some US court guidelines would prefer the professional interpreter ask the speaker to rephrase rather than the professional interpreter doing it (Berk-Seligson, 2017).

PIs also must manage role boundaries (Rosenberg et al., 2008). They are often the only bilingual in the room, so other participants might pressure them to step out of the role (Almeida & Nordin, 2017).

A strong sense of professional ethics helps PIs handle these situations. For example, by using first person and proper introductions to remind everyone of the professional interpreter's function (Hsieh, 2006).

Role boundary issues are especially common in community and medical settings. Clients or providers may not be used to working with PIs and might treat them as co-interviewers or aides (Hsieh, 2010).

One useful concept in recent scholarship is “role-space” (Llewellyn-Jones & Lee, 2014). It frames the professional interpreter's role not as a single point (“neutral” or “advocate”) but as a multi-dimensional space.

Role-space includes factors like presentation of self, degree of involvement, and alignment with participants. PIs dynamically position themselves within this space depending on context (Llewellyn-Jones & Lee, 2014).

4.2.3. Professional identity

As interpreting has professionalized, PIs have come to see themselves (and project themselves) as skilled experts, not “helpmates” (Fernández-Bravo, 2019). Historically, many PIs, especially in community and sign language domains, started as volunteers or informal helpers, like family members interpreting for relatives (Jiménez-Crespo, 2015).

Over time, training programs, accreditation, and professional associations have elevated interpreting to a recognized profession with standards. This shift contributes to

¹³ <https://ausit.org/general-guidelines/>.

identity formation: PIs increasingly identify with a community of practice, with its own jargon, literature, and values (Mikkelsen, 1996).

Pöchhacker notes that the late 20th century brought a professionalization and specialization of PIs. They now often specialize as legal PIs, conference PIs, etc., with distinct career paths (Pöchhacker, 2022).

This can sometimes create tension. For example, conference PIs historically had higher status and payment than community PIs, which influenced how each group viewed their role (Tiselius, 2021).

However, there is a growing sense of a unified profession. Given that PIs in all sectors face similar challenges, particularly in relation to evolving technology and working conditions (Buzungu & Hansen, 2020).

PIs also develop identity through training socialization. In professional interpreter training programs, students are often inculcated with certain role philosophies (Lee, 2020).

For instance, court interpreting courses stress the conduit model and the professional interpreter as “officer of the court” serving justice, which becomes part of how trainees self-identify as impartial judicial PIs (Colin & Morris, 1996). In contrast, a program in community interpreting might encourage the idea of the professional interpreter as a cultural mediator, which the student then integrates into their professional self-concept (Wang, 2017).

These identities can evolve with experience. A new professional interpreter might start rigidly neutral but later realize that flexibility is needed, thus reshaping their own understanding of their role (Lang, 2023).

4.2.4. Situational and social context

A professional interpreter’s social identity (gender, ethnicity, language background) can also affect their role performance (Rudvin, 2008). Research by Nakane (2009) in courtrooms, for example, showed that PIs sometimes experience role conflict if they share the ethnicity of a defendant – they may feel community pressure versus professional duty.

Similarly, sign language PIs, who often are in the same community as their deaf clients, historically shifted roles. From “helper” (pre-1970s, when they were advocates and caregivers for deaf individuals) to “communication facilitator” (after professionalization, focusing on conveying message only) to “bicultural mediator” (recognizing that bridging deaf and hearing cultures may require more than literal interpreting) (Cokely & Marschark, 2005).

PIs' social and ethical positioning is a balancing act. They must uphold ethical standards that safeguard fairness and understanding, yet remain aware of the human and contextual factors that make each interpreted event unique (Marangoz, 2023).

The modern professional interpreter's professional identity is one of a reflective practitioner (Bown, 2013). They are expected to be aware of their role and its limits, to make ethical choices in real time, and to continuously justify and improve their practice according to both communicative outcomes and professional norms.

The evolution from "language conduit" to "language mediator" in Interpreting Studies discourse signifies a more mature understanding that PIs have agency and responsibility in communication. Their role is not just linguistic, but also deeply social.

Understanding the historical evolution of the professional interpreter's role and their social, ethical, and professional identity is essential when studying S2ST because it highlights the human values, decision-making processes, and contextual awareness that current technologies aim to replicate or replace. This perspective allows for a critical evaluation of how S2ST systems align with, diverge from, or potentially undermine the nuanced competencies embedded in human interpreting.

Next chapter provides an overview of S2ST, detailing its three core components, and discussing the evolution, challenges, and advancements of each, including the rise of end-to-end models. It emphasizes the interdisciplinary nature of S2ST, technical and ethical challenges, and the ongoing efforts to enhance real-time multilingual communication through neural architectures and sociolinguistic insights.

5. Speech-to-speech translation

S2ST is the most prominent example of the previously mentioned technologies that are designed to replace PIs, mentioned by Braun (2019, p. 271). It integrates three systems, including automatic speech recognition (ASR), machine translation (MT), and text-to-speech synthesis (TTS) (Nakamura et al., 2006).

The process begins with ASR capturing and converting the speaker's input into text in the source language. This text is then translated into the target language using MT, which currently relies on algorithms and NNs trained on vast multilingual datasets. Finally, the TTS system converts the translated text into speech (Jia et al., 2019).

S2ST research and development is a multidisciplinary field. It integrates technologies and methodologies from Linguistics, Computer Science, and AI to enable real-time (or near-real-time) translation of spoken language.

Advances in the areas of ASR, MT and TTS, as well as their integration, are critical for improving the overall performance of S2ST systems. Modern systems often incorporate neural network-based models, particularly those utilizing transformer architectures, which have significantly enhanced translation accuracy and naturalness (Kala et al., 2025).

One of the primary challenges in S2ST is achieving contextually accurate and culturally appropriate translations (Min et al, 2025; Dhawan, 2022). Spoken language is highly dynamic, with variations in tone, pitch, and prosody influencing meaning (Hirschberg, 2002).

Idiomatic expressions and cultural references often do not have direct equivalents in other languages, necessitating sophisticated algorithms capable of capturing contextual nuances (Marais et al., 2020). Addressing these issues requires not only advanced computational models but also the incorporation of linguistic and sociocultural insights into system design.

Another critical aspect of S2ST is the handling of real-time constraints. In many applications, latency has to be minimized to maintain the flow of conversation, which poses technical challenges (Bangalore et al., 2012).

Processing speed is supposed to be balanced with accuracy and naturalness, requiring optimization of algorithms and efficient hardware implementations (Federico et al., 2020). Emerging techniques, such as end-to-end (E2E) S2ST models that bypass intermediate text

representations, have shown promise in reducing latency while maintaining performance (Li et al., 2023).

Ethical considerations also play a significant role in the development and deployment of S2ST systems. Issues such as data privacy, bias in training data, and potential misuse of the technology is supposed to be carefully addressed (Williams et al., 2023; Lin et al., 2024).

For example, inaccuracies in translation could lead to misunderstandings in critical scenarios, such as medical consultations or legal proceedings (Robertson & Díaz, 2022). Developers are expected to prioritize transparency (Karunya et al., 2023), accountability (Marais et al., 2020), and inclusivity in their approaches to ensure that S2ST systems serve diverse populations effectively and equitably (Gupta et al., 2024).

5.1. The three components

5.1.1. Automatic Speech Recognition (ASR)

The process of converting spoken language into text is a cornerstone of speech technology, with applications ranging from virtual assistants to real-time S2ST systems. This process, known as ASR, involves multiple sophisticated methodologies that bridge the gap between acoustic signals and linguistic representations (Amodei et al., 2016).

ASR systems rely on acoustic modeling, language modeling, and decoding algorithms to transcribe spoken input into written text. These models are trained on extensive datasets to recognize patterns in speech, from phonemes to full sentences, enabling seamless interaction between humans and machines (Graves et al., 2013).

Core methodologies in ASR have evolved significantly over time. Early systems used rule-based approaches that relied on manually designed phonetic and grammatical rules, which were rigid and prone to errors (Wester et al., 2000).

Statistical methods, such as Hidden Markov Models (HMMs) combined with Gaussian Mixture Models (GMMs), marked a significant improvement by modeling speech as a probabilistic process. In these models, HMMs represented the temporal structure of speech, while GMMs captured the acoustic variability (Xuan et al., 2001).

More recently, the advent of deep learning has transformed ASR with end-to-end neural architectures that unify acoustic and language modeling within a single framework (Prabhavalkar et al., 2023). E2E ASR models have become state-of-the-art, with architectures

like Connectionist Temporal Classification (CTC), sequence-to-sequence (Seq2Seq) models, and transformers leading the field (Salazar et al., 2019).

These models bypass traditional pipelines by directly mapping audio inputs to text outputs. Transformers, such as Wav2Vec 2.0 and Whisper, excel at learning contextual dependencies across long audio sequences, enabling them to achieve high accuracy even with complex speech patterns (Zuluaga-Gomez et al., 2023; Graham et al., 2024).

Despite these advancements, Hinsvark et al. (2021) discuss ASR challenges when dealing with accents, dialects, noise, and speaker variability. Accents and dialects introduce phonetic and prosodic variations that can deviate from the patterns observed in training data, leading to transcription errors.

For example, an ASR model trained primarily on standard American English may struggle with non-native speakers or regional dialects. Similarly, background noise and environmental distortions degrade the quality of acoustic signals, making it harder for models to distinguish speech from non-speech sounds (Li et al., 2014).

Speaker variability adds another layer of complexity, as individuals differ in pitch, tone, speed, and articulation. Gender, age, and health conditions can also influence speech characteristics, further complicating recognition tasks (Benzeghiba et al., 2007).

To address these challenges, researchers have turned to data augmentation techniques, domain adaptation, and transfer learning. Kheddar et al. (2023) demonstrate that by exposing models to diverse datasets and simulating real-world conditions, ASR systems can become more robust against variations in speech and environment.

According to Sharab et al. (2025), state-of-the-art ASR models also leverage self-supervised learning to improve performance in challenging scenarios. Models like the aforementioned Wav2Vec 2.0 are pre-trained on massive amounts of unlabeled audio data, learning general representations of speech before fine-tuning on labeled datasets (Baevski et al., 2020).

This approach enables models to generalize better across languages and speaking conditions, enhancing their ability to handle diverse inputs. Additionally, multi-task learning and multilingual training strategies have shown promise in creating ASR systems that can adapt to multiple languages and dialects without the need for separate models (Prajwal et al., 2025).

5.1.2. Machine Translation (MT)

Text-based MT plays an essential function in the S2ST pipeline, serving as the intermediary between ASR and TTS systems. Once the spoken input is converted into text by ASR, MT translates it into the target language, enabling cross-lingual communication (Kenny, 2018).

This translation step plays a critical role in preserving the semantic and pragmatic integrity of the original message. If errors occur, they can cascade through the entire S2ST pipeline, degrading the quality of the translated speech (Min et al., 2025).

Early approaches to MT relied on rule-based systems, which used extensive sets of linguistic rules and bilingual dictionaries to perform translations. While these systems were effective in handling predictable grammatical structures, they struggled with idiomatic expressions, contextual subtleties, and linguistic variability (Shiwen & Xiaojing, 2014).

Creating rules for each language pair was a labor-intensive process that made early systems rigid and difficult to scale. That approach was especially impractical for languages with limited standardization or scarce linguistic resources (Wang et al., 2022).

The introduction of statistical MT (SMT) marked a significant shift in MT methodologies. Instead of relying on manually defined rules, SMT systems used probabilistic models to learn translations from parallel corpora (Och & Ney, 2000).

By analysing patterns and co-occurrences in large datasets, SMT systems could produce more flexible and accurate translations. However, they still faced challenges in handling rare word pairs, long-distance dependencies, and syntactic divergences between languages (Deng & Xue, 2017).

Neural MT (NMT) has emerged as the state-of-the-art approach in MT, transforming the S2ST pipeline with its ability to produce fluent, context-aware translations. NMT models leverage deep learning architectures, such as Recurrent Neural Networks (RNNs) and transformers, to process entire sentences as sequences rather than individual words or phrases (Stahlberg, 2020).

This holistic approach allows NMT systems to better capture contextual and semantic relationships, resulting in translations that are more natural and accurate. Models like Google's Transformer and OpenAI's GPT have further advanced NMT by incorporating attention mechanisms, enabling them to focus on relevant parts of the input text during translation (Vaswani et al., 2017).

One of the critical challenges for MT in the S2ST pipeline is the translation of low-resource languages. Many languages lack the extensive parallel corpora needed to train effective models, leading to lower translation quality compared to high-resource languages (Ranathunga et al., 2023).

To address this, researchers are also exploring techniques such as transfer learning, where knowledge from high-resource languages is transferred to low-resource ones, and unsupervised NMT, which learns translation mappings without parallel corpora. Although they are not enough, these methods are essential for making S2ST technology accessible to speakers of underrepresented languages (Zoph et al., 2016).

Domain-specific translations also present a unique set of challenges and opportunities for MT in the S2ST context (Zheng et al., 2024). General-purpose models may struggle with specialized terminology and jargon found in technical, legal, or medical conversations (Hutchins, 2003).

Fine-tuning NMT models on domain-specific corpora can improve their performance in these contexts, but it requires access to high-quality, domain-relevant data. Additionally, hybrid approaches that combine rule-based or statistical techniques with neural models can be used to handle highly specialized or controlled language environments (Costa-Jussa & Fonollosa, 2015).

5.1.3. Text-to-Speech Synthesis (TTS)

TTS systems convert translated text into speech that is both intelligible and natural-sounding. The goal of TTS systems is to synthesize speech that mimics human speech in terms of pronunciation, rhythm, and intonation, enabling effective communication across languages (Pamisetty & Murty, 2023).

Early TTS systems employed concatenative synthesis, which assembled speech by linking together short, pre-recorded audio segments. Alternatively, they utilized formant synthesis, generating speech through mathematical models that simulated the acoustic properties of the human vocal tract (Story, 2019).

While these methods produced intelligible speech, they often lacked the naturalness and expressiveness needed for realistic communication (Dall et al., 2014). With the advent of machine learning, however, TTS models learned to map text to audio directly from large datasets, bypassing the limitations of rule-based methods (Khanam et al., 2022).

Neural TTS (NTTS) systems, such as the WaveNet and Tacotron, have significantly advanced the field by producing high-quality, natural-sounding speech. WaveNet, developed by DeepMind, uses deep NNs to model raw audio waveforms, capturing intricate details of human speech, including natural variations in pitch, tone, and rhythm (House, 2017).

Tacotron, on the other hand, is a Seq2Seq model that generates spectrograms from input text, which are then converted into audio using a vocoder like WaveNet. These models leverage attention mechanisms to align text and speech features, enabling them to handle complex linguistic patterns effectively (Weiss et al., 2021).

Prosody, which refers to the rhythm, stress, and intonation of speech, is a key factor in achieving naturalness in synthesized speech. NTTS systems have made strides in modeling prosody by learning patterns from large and diverse datasets (Pamisetty & Murty, 2023).

For example, Tacotron 2 introduces prosodic variations by conditioning the model on contextual features, allowing it to generate speech that reflects the intended meaning and emotion of the input text (Aziz et al., 2023). This capability is particularly important in S2ST systems, where maintaining the speaker's intent and tone is critical for effective communication.

Emotion and expressiveness further enhance the naturalness of synthesized speech by making it more relatable and human-like. Advances in expressive TTS involve training models on emotionally labeled datasets, enabling them to generate speech with varying levels of excitement, sadness, or neutrality (Zhang et al., 2025).

Techniques like Global Style Tokens (GSTs) and fine-grained control over acoustic features allow TTS systems to produce speech with dynamic emotional variations, adapting to different conversational contexts. This is particularly useful in applications like customer service, virtual assistants, and storytelling (Wang et al., 2018).

Despite these advances, real-time synthesis poses computational challenges, especially in S2ST systems where maintaining low latency is crucial for natural, fluid interaction (Liu et al., 2021). Ensuring rapid processing without sacrificing speech quality remains a key technical hurdle in making these systems more efficient and user-friendly.

5.2. E2E S2ST Models

E2E S2ST models represent a transformative approach to multilingual communication by directly mapping input speech to output speech without relying on intermediate text-based

MT representations. These unified architectures streamline the traditional S2ST pipeline, which typically involves separate stages for ASR, MT, and TTS (Jia et al., 2022).

By eliminating intermediary steps, E2E models reduce the potential for error propagation, improving latency, and simplifying the training process. These advantages make them a promising solution to achieve high quality (Sperber & Paulik, 2020).

Traditional S2ST systems face challenges due to their modular nature, as errors introduced in one stage can propagate through the pipeline, negatively affecting the final output. In contrast, E2E S2ST models directly learn the complex mappings between input and output speech using a single neural network (Biadsy et al., 2019).

This unified approach leverages advances in Seq2Seq models and attention mechanisms, allowing the system to capture both acoustic and linguistic features in a cohesive framework (Jia et al., 2019). By learning these mappings holistically, E2E models offer the potential for more accurate and contextually appropriate translations.

Recent advancements in Seq2Seq models have played a crucial role in the development of E2E S2ST architectures. These models process input sequences (e.g., audio frames) and generate corresponding output sequences (e.g., translated audio) through an encoder-decoder framework (Nguyen et al., 2020).

The encoder captures the input speech features and compresses them into a latent representation, while the decoder generates the output speech in the target language. Attention mechanisms further enhance this process by dynamically aligning input and output sequences, ensuring that relevant information is emphasized during translation (Tang et al., 2023).

Transformer-based architectures, such as the S2S Transformer, have emerged as a key innovation in E2E S2ST systems. Unlike traditional Seq2Seq models that rely on RNNs, transformers use self-attention mechanisms to process entire sequences in parallel (Wang et al., 2023).

This capability allows transformers to handle long-range dependencies more effectively, making them particularly well-suited for translating complex sentences and maintaining context across extended utterances. Additionally, transformers enable scalability and faster training, making them a cornerstone of modern E2E S2ST research (Zhou et al., 2024).

One notable advancement in E2E S2ST models is the use of speech representations that bridge the acoustic and linguistic domains. Techniques such as speech embeddings and self-supervised pretraining (e.g., Wav2Vec 2.0) have enabled models to extract rich feature

representations from raw audio, facilitating more accurate and efficient translation (Fukuda et al., 2023).

Unsupervised and semi-supervised approaches have shown potential in addressing the scarcity of parallel speech datasets. This scarcity has long been a significant obstacle to the development of E2E S2ST systems (Wang et al., 2022).

Despite their promise, those systems face several challenges, particularly in achieving robust performance across diverse languages and speakers. The lack of large-scale, high-quality parallel datasets for many language pairs limits the ability of these models to generalize effectively (Gris et al., 2024).

Additionally, generating natural-sounding output speech with appropriate prosody and emotion remains a complex task, requiring sophisticated speech synthesis techniques. Computational requirements for training and inference also pose challenges, particularly for real-time applications where low latency is essential (Pamisetty & Murty, 2023).

PART II

6. Professional interpreter's competence

The corpus of academic publications on interpreting that we attached to ChatGPT-o1 contains 264 files published in English, as shown in Table 3. The publication dates cover the period from 2004 to 2024. The data occupy 435 MB of computer storage; we combined them into a single PDF using Adobe Acrobat Pro.

The types of publications recovered include PhD dissertations, Master's theses and papers. In total, 215 files had their metadata edited in Mendeley, including title, authors, publication year and DOIs. Metadata editing took 135 days to complete.

Table 3 - Interpreting corpus size.

Corpus numbers		Publication types	
Files	264	PhD dissertations	13
Period	2004 - 2024	Master's theses	7
Size	435 MB	Papers	244

Source: the author.

The prompt given to ChatGPT-o1 was: “Based on the academic publications attached to this chat, what knowledge and skills build the competence of a professional interpreter?”. The response given by the tool was a 131-word report that we discuss in this chapter.

6.1. Linguistic and communicative competence

The first set of knowledge and skills retrieved by ChatGPT-o1 in the corpus of interpreting academic publications is “linguistic and communicative competence”. Littig & Pöchhacker (2014) mention this expression, without, however, specifying what it is about.

Other authors mention these competencies separately (Zinukova, 2016). However, some authors refer to “linguistic competence” as “language competence” (Setton & Dawrant, 2016), “language knowledge” (Jones, 2014) or “linguistic skill” (Macnamara et al., 2014).

Table 4 shows the main components of linguistic and communicative competence developed by PIs and retrieved by ChatGPT in the interpreting corpus. Evidently, each professional interpreter has a more or less diversified domain of these skills.

Table 4 - Linguistic and communicative competence.

1. Linguistic and communicative competence

- Near-native proficiency in at least two languages (source and target).
- Deep understanding of syntax, semantics, and pragmatics in both languages.
- Mastery of idiomatic expressions, slang, and dialectal variations.
- Awareness of false cognates and ambiguities that could cause misinterpretations.

Source: ChatGPT/Interpreting Corpus.

PIs are expected to possess near-native proficiency in at least two languages to ensure accurate and effective communication (Loiseau & Delgado-Luchner, 2021). This high level of proficiency enables PIs to comprehend and convey messages, capturing the nuances and intentions of the original discourse.

A deep understanding of syntax, semantics, and pragmatics in both languages is essential for PIs (Defrancq & Collard, 2020). Syntax involves the arrangement of words to create well-formed sentences, semantics pertains to the meaning conveyed, and pragmatics focuses on the use of language in context (Kroeger, 2023).

PIs must also be adept with idiomatic expressions, slang, and dialectal variations (Sun et al., 2022). These elements often carry cultural significance and can differ widely between regions and communities.

A professional interpreter's ability to recognize and accurately translate such expressions ensures that the intended message resonates with the target audience, preserving the original's tone and intent. This skill is particularly crucial when working with diverse populations where language usage can vary significantly.

Awareness of false cognates and potential ambiguities is another critical competency for PIs (Gieshoff, 2017). False cognates are words that appear similar in two languages but have different meanings, which can lead to significant misunderstandings if misinterpreted (Marecka et al., 2021).

By identifying and addressing these linguistic pitfalls, PIs maintain the integrity of the communication. It is a way to ensure that messages are conveyed accurately and effectively across language barriers.

6.2. Cognitive and Memory Skills

The second set of knowledge and skills retrieved by ChatGPT in the corpus is “cognitive and memory skills”. Čeňková et al. (2014) call it “cognitive abilities” and no author in the corpus mentions “memory skills”, which makes us think that ChatGPT has inferred this from semantic approximations.

PIs require cognitive and memory skills to perform effectively in demanding communication scenarios, as shown in Table 5. One foundational skill is active listening (Setton & Dawrant, 2016), which involves fully engaging with the speaker, comprehending the intended meaning, and analysing both explicit and implicit messages (Tennant et al., 2017).

Table 5 - Cognitive and memory skills.

2. Cognitive and Memory Skills

- Active listening
- Short-term memory retention
- Rapid information processing
- Multitasking

Source: ChatGPT/Interpreting Corpus.

Short-term memory retention is another critical cognitive skill, as PIs must temporarily hold significant chunks of information while processing the meaning and translating accurately (Russell & Takeda, 2015). Effective short-term memory allows PIs to manage information flow, especially during simultaneous interpretation, where retention of recent dialogue is necessary to ensure continuity and accuracy in the conveyed message (Babcock et al., 2017).

The ability to process information quickly is essential for interpretation. PIs are expected to quickly analyse, translate, and articulate information in real-time scenarios (Setton & Dawrant, 2016).

Quickly processing information involves the ability to swiftly identify key information, interpret meaning accurately, and produce coherent translations immediately. Such rapid processing supports PIs in adapting to various linguistic structures and contextual complexities efficiently (Chen, 2017).

PIs are expected to excel at multitasking, simultaneously engaging in listening, analysing, translating, and speaking activities. The ability to multitask effectively ensures that PIs maintain accuracy, fluency, and timing, even under intense cognitive loads (Bartłomiejczyk & Stachowiak-Szymczak, 2021).

6.3. Cultural and contextual awareness

The third set of knowledge and skills retrieved by ChatGPT in the corpus is “cultural and contextual awareness”, as shown in Table 6. Setton & Dawrant (2016) mention “contextual knowlegde” and Littig & Pöchhacker (2014) mentions “cultural competence”.

Table 6 - Cultural and contextual awareness.

3. Cultural and contextual awareness

- Knowledge of social norms, taboos, and politeness strategies in both cultures.
- Understanding historical and political contexts that influence discourse.
- Ability to adapt tone and register to match speakers and audiences.

Source: ChatGPT/Interpreting Corpus.

Cultural and contextual awareness are indispensable for PIs, as they try to bridge linguistic divides and deal with intricate social norms, taboos, and politeness strategies inherent in diverse cultures (Munane, 2014). Understanding of these elements enables PIs to convey messages that are not only accurate but also culturally resonant, thereby preventing potential misunderstandings.

PIs are expected to grasp the historical and political contexts that shape discourse within the cultures they work with (Baker & Diriker, 2019). Awareness of historical events, political dynamics, and societal values allows PIs to comprehend underlying references and connotations in conversations, which might otherwise be lost or misinterpreted (Valdés, 2014).

Additionally, the ability to adapt tone and register to align with both the speaker and the audience is vital (Setton & Dawrant, 2016). Different cultures have varying expectations regarding formality, directness, and the use of honorifics (Brown, 2015).

6.4. Subject-matter competence

The fourth set of knowledge and skills retrieved by ChatGPT in the corpus is “subject-matter competence”, shown in Table 7. Xu (2018) mentions “subject-matter knowledge”, and Setton & Dawrant (2016) mention “domain-knowledge”.

Table 7 - Subject-matter knowledge.

4. Subject-matter competence • Domain-specific knowledge • Terminology management
--

Source: ChatGPT/Interpreting Corpus.

Subject-matter competence in interpreting refers to a professional interpreter’s in-depth understanding of the specific field or industry related to the content being interpreted (Ruffolo, 2023). Understanding of specific fields, such as law, medicine, or engineering, enables PIs to grasp intricate concepts and terminologies, ensuring precise and contextually appropriate interpretation (Alarcón, 2025).

Effective terminology management is crucial in maintaining consistency and clarity during interpretation. By systematically organizing and standardizing specialized terms, PIs can ensure that the correct terminology is used consistently, reducing the risk of misunderstandings and enhancing communication effectiveness (Xu, 2015).

PIs with domain-specific knowledge can anticipate subject matter complexities, facilitating smoother communication between parties. Their competence allows them to navigate discussions, ensuring that all conveyed information is both accurate and culturally appropriate (Jutorán, 2023).

Utilizing terminology management systems further supports PIs in handling specialized vocabularies efficiently. These systems provide centralized repositories for terms, definitions,

and usage guidelines, aiding PIs in delivering consistent and high-quality interpretations across various contexts (Andrews, 2025).

6.5. Ethical and professional conduct

The last set of knowledge and skills retrieved by ChatGPT in the corpus is “ethical and professional conduct”, shown in Table 8. Berk-Seligson (2017) mentions “ethical behavior and professional conduct”, and Baker & Diriker (2019) mention “ethical dimensions of interpreting”.

Table 8 - Ethical and professional conduct.

5. Ethical and professional conduct ● Confidentiality ● Neutrality ● Composure under pressure

Source: ChatGPT/Interpreting Corpus.

Ethical and professional conduct is paramount in interpreting, with confidentiality being a foundational principle (Carballido et al., 2024). PIs are obligated to keep all information acquired during their private session, ensuring that sensitive details remain secure and fostering trust between all parties involved (Edwards; Temple; Alexander, 2005).

Neutrality is another critical aspect, requiring PIs to remain impartial and unbiased throughout their work. By refraining from inserting personal opinions or biases, PIs ensure that communication is accurate and equitable, allowing parties to make informed decisions based on unaltered information (Ayan, 2020).

Maintaining composure under pressure is essential for PIs, especially in high-stakes or emotionally charged environments (Chilingaryan et al., 2016). The ability to stay calm and focused enables PIs to deliver accurate interpretations, uphold professional standards, and effectively manage the complexities inherent in their role (Burns, 2010).

7. Speech-to-speech development stages

The corpus of academic publications on S2ST attached to ChatGPT contains 953 files, as shown in Table 9. The publication dates cover the period from 2004 to 2024. The data occupy 954 MB of computer storage; we combined them into a single PDF using Adobe Acrobat Pro.

The types of publications recovered include PhD dissertations, Master's theses, and papers. In total, 953 files had their metadata edited in Mendeley, including title, authors, publication year and DOIs. Metadata editing took 251 days to complete.

Table 9 - S2ST Corpus size.

Corpus numbers		Publication types	
Files	953	PhD dissertations	31
Period	2004 - 2024	Master's thesis	57
Size	954 MB	Papers	865

Source: the author.

The prompt we gave to ChatGPT-o1 was: “Based on the academic publications attached to this chat, what are the development stages of a speech-to-speech translation system?”. The response given by the tool was a 75-word report that we discuss in this chapter.

7.1. Data collection and preparation

The first stage in S2ST development retrieved by ChatGPT is data collection and preparation. The process involves creating speech corpora, parallel text corpora and audio-text alignment, as shown in Table 10.

Table 10 - Data collection & Preparation.

1.	Data	collection	and	Preparation
	• Speech corpora • Parallel text corpora • Audio-text alignment			

Source: ChatGPT/S2ST Corpus.

Speech corpora consist of audio samples reflecting the full range of linguistic diversity. They can include different accents, dialects, speech styles (formal, informal, spontaneous) (Kumar et al., 2014), and acoustic settings (quiet, noisy, indoor, outdoor) (Veluri et al., 2024).

Parallel text corpora are equally important for the translation component, requiring extensive datasets of aligned text pairs in both the source and target languages. Precise alignment of these corpora directly enhances translation accuracy and fluency, influencing the system's ability to deliver meaningful and contextually appropriate translations (Uliniansyah et al., 2017).

Audio-text alignment, where each speech segment is accurately matched to its textual transcription or translation, is critical to the learning process. Proper alignment ensures that the model correctly interprets spoken language and matches it with corresponding textual data (Alastruey et al., 2023).

7.2. ASR model development

The second stage in S2ST development retrieved by ChatGPT is ASR model development. This stage involves feature extraction, neural network training and post-processing, as shown in Table 11.

Table 11 - ASR model development.

2. ASR model development

- Feature extraction
- Neural network training
- Post-processing

Source: ChatGPT/S2ST Corpus.

The first step is feature extraction, where raw audio signals are transformed into machine-readable formats, such as Mel-frequency Cepstral Coefficients (MFCCs) or spectrograms (Lei, 2024; Tjandra et al., 2019). These processed features capture important acoustic properties, facilitating analysis by NNs.

The next step is neural network training, commonly utilizing advanced Deep Neural Networks (DNNs) like Transformers, RNNs, or hybrid approaches (Kano et al., 2018). These models are trained on extensive datasets to accurately map audio features to corresponding textual output.

Finally, post-processing steps significantly enhance the model's output quality (Mujadia & Sharma, 2023). Techniques such as language modeling are employed to correct misrecognized or out-of-vocabulary words, while punctuation insertion ensures the readability and naturalness of the transcribed text (Wu et al., 2024).

7.3. MT model development

The third stage in S2ST development retrieved by ChatGPT is text translation model development. The process involves preprocessing and tokenization, NMT and domain adaptation, as shown in Table 12.

Table 12 - Text translation model development.

3. MT model development

- Preprocessing & tokenization
- Neural machine translation (NMT)
- Domain adaptation

Source: ChatGPT/S2ST Corpus.

Preprocessing and tokenization requires meticulous handling of textual data. This includes normalizing text by converting it to lowercase, removing punctuation, and expanding contractions to maintain consistency (Camacho-Collados & Pilehvar, 2017; Schmidt et al., 2024).

Tokenization then segments the text into manageable units, such as words or subwords, facilitating effective processing by translation models. Advanced tokenization techniques, like Byte Pair Encoding (BPE), are particularly beneficial as they balance vocabulary size and model complexity, effectively managing rare and compound words (Bostrom & Durrett, 2020).

Following preprocessing, the implementation of NMT models marks a significant advancement over traditional statistical methods (Stahlberg, 2020). NMT models, especially those based on the Transformer architecture, excel in capturing long-range dependencies within text, leading to more fluent and accurate translations (Araabi & Monz, 2020).

Unlike earlier models that processed text sequentially, Transformers utilize self-attention mechanisms to evaluate entire sentences simultaneously, enhancing both efficiency and contextual understanding. The ability to process in parallel is essential for faster training and higher quality translations (Raganato & Tiedemann, 2018).

To further refine translation models for specific applications, domain adaptation is employed. This process involves fine-tuning pre-trained NMT models on domain-specific datasets, enabling the model to grasp and accurately translate terminology unique to particular fields (Saunders, 2022).

Techniques such as transfer learning and the integration of domain-specific data during training have proven effective in this context. For instance, incorporating a k-nearest-neighbor (kNN) classifier over an external datastore can enhance translation accuracy without necessitating complete model retraining (Gao et al., 2024).

7.4. TTS model development

The fourth stage in S2ST development retrieved by ChatGPT is TTS model development. The process involves text normalization, acoustic modelling and vocoder implementation, as shown in Table 13.

Table 13 - TTS model development.

4. TTS model development

- Text normalization
- Acoustic modelling
- Vocoder implementation

Source: ChatGPT/S2ST Corpus.

The initial step to TTS model development is text normalization. This process transforms raw translated text into a format suitable for speech synthesis.

The TTS model addresses the conversion of numbers, abbreviations, dates, and other non-standard words into their full, spoken equivalents. Its objective is to ensure clarity and naturalness in the synthesized speech (Dang & Phan, 2022).

For instance, the numeral “2025” is normalized to “two thousand twenty-five,” and the abbreviation “Dr.” becomes “Doctor.” Effective text normalization enhances the intelligibility of the output, laying a solid foundation for subsequent synthesis stages (Zhang et al., 2019).

Following normalization, the acoustic model plays a crucial role by converting the prepared text into a representation of acoustic features, such as spectrograms. These features encapsulate essential aspects of the speech signal, including frequency and amplitude variations over time (Shen et al., 2018).

Modern acoustic models often employ deep learning architectures, such as recurrent or transformer-based NNs, to predict these features with high accuracy. By effectively modeling the complex relationships between linguistic input and acoustic output, these models ensure that the synthesized speech closely mirrors natural human speech patterns (Li et al., 2022).

The final stage involves the vocoder, which converts the acoustic features generated by the acoustic model into an audio waveform. Contemporary vocoders, like WaveNet and WaveGlow, utilize advanced neural network techniques to produce high-fidelity, natural-sounding speech (Prenger et al., 2019).

These models have significantly advanced the quality of synthesized speech, reducing the robotic or artificial qualities that characterized earlier systems. By reconstructing the waveform from acoustic features, vocoders ensure that the final audio output is both intelligible and engaging, effectively delivering the translated content in a human-like manner (Gu et al., 2024).

7.5. Integration and real-time system design

The fifth stage in S2ST development retrieved by ChatGPT is integration and real-time system design. The process involves latency reduction, APIs and interfaces and edge or cloud deployment, as shown in Table 14.

Table 14 - Integration and real-time system design.

5. Integration and real-time system design

- Latency reduction
- APIs and interfaces
- Edge or cloud deployment

Source: ChatGPT/S2ST Corpus.

Latency is a critical factor in S2ST systems that operate in real-time or near-real-time. To achieve minimal delays, it is essential to decompose the translation pipeline into streaming components (Wadhawan, 2023).

Employing more compact and efficient models can further reduce processing times, ensuring swift and seamless translations. For instance, Google's SimulTron model exemplifies an on-device, real-time S2ST system optimized for low latency (Agranovich et al., 2024).

Integrating ASR, NMT, and TTS modules into a cohesive workflow necessitates well-designed APIs and interfaces. This integration ensures smooth data flow between components, reducing overhead and potential bottlenecks (Hayashi et al., 2020).

Utilizing standardized protocols and efficient data serialization formats can enhance communication efficiency. For example, implementing adaptive buffering techniques in TTS pipelines can mitigate latency, providing smoother playback in conversational AI applications (Ma et al., 2019).

Deciding between edge or cloud deployment involves balancing factors such as speed, cost, and privacy. Edge computing processes data locally on devices, offering reduced latency and enhanced privacy by minimizing data transmission (Khan et al., 2019).

In contrast, cloud computing provides scalable resources but may introduce latency due to data transmission times. For applications requiring immediate responses and data security, edge deployment is advantageous (Sunyaev & Sunyaev, 2020).

Tasks demanding extensive computational power and scalability may benefit from cloud solutions. Understanding the specific requirements of a S2ST application is crucial in making an informed deployment choice.

7.6. Evaluation and iteration

The sixth stage in S2ST development retrieved by ChatGPT is evaluation and iteration. This stage includes metric-based and/or human evaluation, error analysis and system updates, as shown in Table 15.

Table 15 - Evaluation and iteration.

6. Evaluation and iteration

- Metric-based
- Human evaluation
- Error analysis
- System updates

Source: ChatGPT/S2ST Corpus.

Evaluating S2ST systems necessitates the use of specialized metrics tailored to each component of the pipeline. For ASR, Word Error Rate (WER) is the standard metric, quantifying the percentage of words incorrectly transcribed (von Neumann et al., 2023).

In machine translation, metrics like BLEU, ROUGE, and METEOR assess the quality of translated text by comparing it to reference translations, focusing on aspects such as n-gram overlap and recall (Mathur et al., 2020; Chatoui & Ata, 2021; Saadany & Orasan, 2021). TTS synthesis, the Mean Opinion Score (MOS) is employed, wherein listeners rate the naturalness and intelligibility of synthesized speech on a subjective scale (Streijl et al., 2016).

Despite the utility of automated metrics, human evaluation remains indispensable in the assessment of S2ST systems. Human judges can discern subtleties such as contextual appropriateness, fluency, and the preservation of prosody—elements that automated metrics might overlook (Huang et al., 2023).

For instance, studies have shown that automatic metrics may not correlate strongly with human judgments, especially in the context of dialectal variations and expressive speech (Hamed et al., 2023). Therefore, incorporating human feedback ensures a comprehensive evaluation, capturing aspects of translation and synthesis quality that are critical for user satisfaction.

To enhance system performance, conducting thorough error analysis is crucial. This involves scrutinizing segments where the system underperforms, such as those containing domain-specific terminology, background noise, or diverse accents (Soni, 2025).

Identifying these challenging areas allows for targeted improvements, such as adjusting model hyperparameters, optimizing architectures, or augmenting training datasets with representative examples (Bahar et al., 2021). Implementing a cycle of evaluation, analysis, and refinement is essential for the iterative enhancement of S2ST systems, ensuring they meet the demands of real-world applications (Chen et al., 2023).

7.7. Maintenance and continuous improvement

The last stage in S2ST development retrieved by ChatGPT is maintenance and continuous improvement. This stage includes user feedback loop, model monitoring, and versioning and deployment, as Table 16 shows.

Table 16 - Maintenance and continuous improvement.

7. Maintenance and continuous improvement

- User feedback loop
- Model monitoring
- Versioning and deployment

Source: ChatGPT/S2ST Corpus.

Implementing a robust user feedback loop is a critical component of this process. By systematically collecting real-world usage data and user feedback, developers can identify errors and areas needing enhancement (Weinstein, 2002; and our own work, Freitas, 2022).

Addressing these issues through retraining or fine-tuning the models ensures that the system evolves in alignment with user needs and language dynamics. For instance, incorporating user corrections can significantly improve translation accuracy over time (Yun et al., 2014). This approach, however, entails risks. Malicious users may force feedback, for instance, to teach the model to swear or produce offensive language.

Continuous model monitoring is vital, as language is inherently dynamic, with usage patterns and expressions evolving over time. Regularly assessing the system's performance helps detect model drift— where the system's accuracy degrades due to changes in language use or the emergence of new terms (Gupta et al., 2023).

The reliability and accuracy of a S2ST system can be maintained by updating training data or adjusting model parameters. This is achieved through early detection of such drifts which allows for timely intervention (Fuccellaro, 2024).

Effective versioning and deployment strategies are fundamental to managing updates and maintaining system stability. Keeping detailed records of model versions facilitates tracking changes and understanding their impacts on performance (Ding et al., 2024).

Implementing rollback strategies ensures that if a new model version introduces unforeseen issues, the system can revert to a previous stable state, minimizing disruptions for users (Paez, 2018). Maintaining a balance between deploying new features and ensuring system reliability is crucial for sustaining user trust and satisfaction, as we concluded in our Master's thesis (Freitas, 2022).

8. Potential contributions of PIs to S2ST

Based on the data presented in chapters 6 and 7, obtained from the two corpora compiled for this study, ChatGPT-o1 rendered the potential contributions of PIs to S2ST. According to the AI model, data indicates that those systems can be profoundly enhanced through insights derived from PIs, particularly in user experience (UX), linguistic accuracy, and ethical integrity.

PIs' real-time communication skills can inspire sophisticated error-handling mechanisms, such as clarification prompts and confidence indicators, which mimic how humans negotiate misunderstandings. Their involvement in interface design can lead to features like multimodal input, intuitive turn-taking cues, and culturally inclusive visual elements that accommodate users of different backgrounds, literacy levels, and abilities.

Crucially, PIs can help ensure that S2ST systems do not merely interpret words, but search for preserving tone, discourse coherence, and conversational flow—preventing socially awkward or misleading outputs. Their cultural competence can also guide systems in avoiding offensive phrases, adapting to dialectal nuances, and embedding context-aware prosody and emotion into synthetic voices, yielding speech that is both accurate and naturally expressive.

Ethical considerations and real-world deployment strategies further demonstrate the irreplaceable value of PIs in developing robust, inclusive S2ST systems. PIs can serve as ethical auditors, identifying biases in gender, race, or profession embedded in machine output, and recommending mitigation techniques that foster neutrality and fairness across all languages.

PIs' input can ensure privacy-preserving features, such as local processing and secure data handling, especially in sensitive domains like healthcare or legal contexts. In field applications, PIs can help fine-tuning systems for domain-specific use, curate critical terminology, and stress-test AI through live event trials, shadowing, and iterative feedback.

8.1. Design and User Experience (UX)

PIs are adept at managing miscommunication in real time, which can inspire better error handling in S2ST systems. In natural conversations, people continuously ground their

understanding by exchanging signals (e.g. saying “mm-hmm” or rephrasing) to prevent or fix errors (Bulyko et al., 2005).

Notably, humans faced with an unclear utterance rarely say “please repeat” verbatim; instead, they employ strategies like paraphrasing the part they understood or asking a pointed clarification (Testoni & Fernández, 2024). By studying these behaviors, designers can implement similar strategies in S2ST UX.

For instance, an interpreting app could display a partial transcription for user confirmation or offer a gentle clarification prompt when uncertain, rather than defaulting to a generic error message. This approach mirrors how a professional interpreter actively seeks confirmation to ensure accuracy and maintain conversational flow (Ono & Yang, 2024).

Research prototypes of interactive S2ST have tested features like voice commands and on-screen back-translations. Saying “scratch that”, for example, to delete a misrecognized phrase. These tools allow users to verify what the system heard, enhancing transparency and control during interactions (Seligman & Dillinger, 2006).

Involving experienced PIs in the S2ST design phase can help identifying which error-recovery cues are most intuitive. Their insight can lead to error-handling mechanisms that feel more natural. For instance, having the system proactively repeat its understanding of a tricky name or number for verification, akin to a professional interpreter double-checking critical information.

8.1.1. Improving user accessibility and inclusivity

PIs have long been enablers of communication for diverse populations (Okoniewska, 2022), and their competence can guide S2ST systems to be accessible and inclusive. One key lesson is that “one size fits all” interfaces often fail for users of different ages, literacy levels, or abilities (Medhi et al., 2011).

PIs know that not everyone can speak clearly or interact with technology easily. For example, in healthcare settings, some patients are not comfortable with complex devices, and ambient noise or accent variability can make purely voice-driven systems unreliable (Seligman & Dillinger, 2006).

To address this, developers (with PIs’ input) can add multimodal interaction options. For instance, a S2ST system might also allow input via typing or selecting illustrated phrases,

which makes it usable for people who are hard of hearing, less tech-savvy, or in noisy environments.

PIs and translators emphasize inclusion in language support: which languages or dialects are offered can determine who benefits. They advocate for including regional dialects and minority languages, not just major languages, to avoid marginalizing any group (Bowker, 2021).

In practice, involving PIs in UX design can lead to features like simplified modes for novice users and culturally appropriate imagery/icons. The result is S2ST interfaces that accommodate a broad range of users, ensuring that the technology bridges communities instead of excluding them.

8.1.2. Refining interface design for better human-machine interaction

PIs can contribute significantly to the interface and interaction design of S2ST systems. An expert professional interpreter can intuitively understand what feedback or controls a user might need during a bilingual conversation (Dong & Li, 2020).

Early prototyping methods such as Wizard-of-Oz experiments have leveraged PIs as stand-ins for the system to trial different interface ideas. In these studies, a concealed professional interpreter produces interpretations while users interact with what appears to be an automated system – allowing designers to observe user expectations before the real system is built (Schlögl et al., 2015).

Insights from such collaborations can lead to UI refinements. For example, PIs can suggest that showing the source speech as text (for transparency) and a confidence indicator can help users gauge when to trust the machine’s output or repeat themselves.

They can also recommend efficient turn-taking signals, such as a sound or visual cue when the system is “listening” versus “speaking”, to mimic the rhythm a professional interpreter maintains with speakers (Dose, 2020). These design elements make interactions smoother.

By tapping PIs’ understanding of conversation flow, S2ST developers can refine interfaces for seamless human-machine interaction. According to ChatGPT-o1, the end goal is ambitious: a S2ST system that behaves like a considerate professional interpreter, managing turn-taking, providing contextual cues, and handling interruptions gracefully, which can result in a more fluid and user-friendly experience.

8.2. Linguistic and cultural adaptation

8.2.1. Ensuring pragmatic accuracy and discourse-level coherence

Beyond literal translation, PIs excel at conveying the pragmatic intent behind words (e.g. tone, formality, implied meaning) and maintaining coherence across a conversation (Onyshchak et al., 2023). These abilities are critical for S2ST systems to emulate.

Research has found that purely correct but pragmatically awkward interpretations can cause serious miscommunication. An otherwise grammatical sentence delivered in an unsuitable tone or context can mislead listeners, who might interpret the odd phrasing as intentional rudeness or confusion (Horiguchi, 1997).

For example, if an S2ST system ignores politeness conventions (like omitting a polite form of “you” in Japanese or Spanish), the result may sound abrupt or disrespectful (Jakobsson, 2016). PIs can guide systems to avoid such pitfalls by injecting context-appropriate phrasing and politeness markers.

PIs also help with discourse-level consistency, ensuring that pronouns, references, or formality remain consistent throughout a dialogue (Zhan, 2012). Machines often interpret sentence by sentence and can lose thread of who “he” or “she” refers to, or whether the tone should stay formal (Bawden, 2018).

By reviewing transcripts, PIs can identify these lapses and suggest improvements (for instance, if a person has been referred to as “Dr. Smith” initially, later references should not suddenly switch to “the man”). As Oddmund Braaten, CEO of an Interprefy¹⁴, noted only skilled linguists can handle the subtle context, tone, humor, and idioms in speech – areas where AI still struggles.

8.2.2. Adapting interpretation for different cultural contexts and dialects

Language is deeply tied to culture. An expression that is benign in one culture might be offensive or confusing in another (Gao, 2002).

¹⁴ Interprefy, a platform for remote interpreting, reported supporting over 50,000 multilingual meetings including high-profile events like press conferences and even an astronaut interview. <https://www.interprefy.com/>

PIs are trained to navigate these cultural nuances, making them invaluable for adapting S2ST outputs to be culturally appropriate. They can provide guidance on local idioms, gestures, or taboo terms to avoid in target languages.

One dramatic example of cultural context gone wrong was in 2017: Facebook’s S2ST rendered an Arabic greeting “good morning” as “attack them” in Hebrew, leading to an innocent man’s arrest (Al Sharou & Specia, 2022). A professional interpreter or cultural consultant could instantly have caught this error, understanding the sociopolitical context and the idiomatic meaning.

To prevent such incidents, PIs can help curate interpretation choices for sensitive contexts. For instance, ensuring that political slogans or religious phrases are handled carefully by S2ST systems.

They can also advise on dialectal differences. A speech translator might perform well in standard Spanish, but without dialect adaptation it could stumble on Latin American versus Castilian idioms or misinterpret region-specific vocabulary.

By leveraging PIs from different locales, developers can adapt models or offer locale settings (e.g. Mexican Spanish vs. Spain Spanish) so the interpretations resonate with the intended audience. PIs effectively act as cultural liaisons during development, pointing out where a direct translation might violate cultural norms or simply confuse listeners.

Recent developments show a push in this direction: some AI translation platforms now integrate cultural tone settings, and predictions for 2025 foresee significant improvement in capturing idiomatic and cultural context, thanks in part to curated cultural databases (Kudo, 2025). In practice, this means an S2ST system could automatically substitute a culturally equivalent phrase rather than doing a literal interpretation.

By embedding cultural intelligence sourced from PIs, S2ST systems can provide translations that make sense locally, preserving intent and respect across different cultures and dialects. These aspects, however, are difficult for the machine to assimilate, as we will discuss in Section 9.

8.2.3. Improving prosody and speech naturalness

Prosody is a huge factor in how natural and meaningful an interpreted utterance sounds (Tsiamas et al., 2024). Machines typically focus on what is being said (the words), but PIs also pay attention to how it is said.

They modulate their voice to convey emphasis, emotion, or sentence type (raising tone at the end of a question, for example) (Gastó Jiménez, 2018). In S2ST research and development, PIs can contribute by helping systems better reproduce natural prosody in the target speech.

This might involve annotating emphasis and emotion in the training data or tuning the TTS output. It has been observed that emotion and prosody are crucial considerations for truly effective speech translation (Brazier & Rouas, 2024).

If a speaker is excited or angry, a monotone interpretation loses important information (Mozziconacci & Hermes, 1999). PIs, skilled in capturing tone, can help design S2ST systems to detect such cues (perhaps via acoustic features) and reflect them in the synthesized voice.

Recent E2E models like Meta’s Direct S2ST have even tried to carry over the speaker’s vocal characteristics and intonation into the other language, so the listener hears a voice with similar emotion (Lee et al., 2021). Studies indicate that improving prosodic modeling will also solve multicultural issues, meaning it helps listeners of different languages perceive the speech as intended (Duret et al., 2023).

For example, polite or formal intonation patterns differ by culture (DeLongová, 2010), and a professional interpreter is expected to know when to soften delivery or adopt a certain pitch pattern to match expectations. These insights can inform the TTS style guidelines for each language.

Already, technology is moving this way. Developers are working on preserving the original speaker’s pitch and speaking rate in the interpreted audio (sometimes via voice cloning) to make the experience more authentic (Kudo, 2025).

By consulting voice professionals and PIs, S2ST systems can be tuned for naturalness, using appropriate pauses, emphasizing key words, and sounding less robotic. Ultimately, this human-influenced refinement of prosody ensures that the translated speech not only conveys the right meaning but also feels like genuine, engaging speech to the listener.

8.3. Ethical Considerations

8.3.1. *Identifying and mitigating biases in automated translations*

S2ST systems can inadvertently amplify biases present in their training data, a serious ethical concern (Lloyd, 2018). PIs, bound by codes of ethics, are trained to avoid biased language and can play a role in auditing and correcting such issues in S2ST outputs.

Biases can manifest in gender (e.g. always translating a doctor as “he”), in professions, in racial or cultural references, etc. For instance, it was found that some NMT engines translated sentences in ways that reinforced gender stereotypes, such as rendering “the nurse” with a female pronoun and “the engineer” with a male pronoun even when the source language was gender-neutral (Propio, 2024).

PIs, who strive to be accurate and impartial, can help flag these distortions. They might review system outputs and highlight biased patterns (perhaps the system consistently uses male terms for leadership roles, or uses informal speech for certain ethnic names).

With this feedback, developers can implement bias mitigation. For example, balancing the training data or adding rules to preserve neutrality.

A mitigation strategy inspired by human insight involves programming the system to present both gender options when the input is ambiguous, as some translation services have begun to implement (Fucci et al., 2023). Alternatively, the system can randomly vary gender in sample sentences to prevent reinforcing a default or stereotypical choice.

PIs also contribute to inclusive terminology. They can advise the system to use respectful and current terms for groups (e.g. avoiding outdated or offensive terms that may be in training data). Some recent efforts, like Propio’s approach to MT, explicitly incorporate steps to produce inclusive results that reflect diversity (Propio, 2024).

This involves understanding how biases arise from data and adding layers of human oversight to correct them. In summary, PIs can act as a check against the blind spots, helping S2ST developers identify problematic translations and adjust the system to be fair and unbiased.

8.3.2. Ensuring fairness and inclusion in multilingual settings

Fairness in S2ST not only refers to avoiding bias within a language, but also providing equitable service across languages and user groups (Barrault et al., 2023). A interpreter might perform excellently for English–Spanish (abundant training data) but poorly for Swahili–French, which means speakers of less-supported languages get inferior service. PIs, especially those working with marginalized languages, can push for these gaps to be addressed by contributing data or evaluating quality in those languages.

Their involvement ensures the project’s scope remains broad: truly “universal” interpretation means no community is left behind. Another aspect is inclusivity in how the system treats different modes of speech.

PIs are sensitive to variations like non-native accents, colloquial speech, or code-switching, which are common in multicultural conversations (Angermeyer, 2010). They can help the system handle these smoothly (for instance, by identifying when a name or phrase does not need interpretation, or when a user switches language mid-sentence).

The goal is to avoid a scenario where the system works only for “standard” speakers but fails for anyone who speaks differently. Including PIs in the testing phase with diverse speakers can uncover such issues.

Moreover, PIs are expected to adhere to principles of impartiality and respect, treating all parties equally (WRHA, 2021). They can impart these values to S2ST design, suggesting features like equal-volume output for all languages (so one language is not literally quieter or less audible than another) and avoiding any pejorative tone.

In live multi-language settings, fairness might also involve turn-taking policies, ensuring the system does not consistently lag more for one language direction than the opposite. Through continual input from PIs, S2ST systems can be tuned to provide a balanced and inclusive experience for all language pairs and speaker communities, reflecting the professional interpreter’s commitment to linguistic equity.

8.3.3. Addressing privacy concerns related to live interpretations

Privacy is a paramount ethical concern, especially as S2ST moves into live settings like business meetings, medical consultations, or legal contexts (Platnick et al., 2024). PIs are usually bound by strict confidentiality, they are expected not to disclose or misuse the sensitive information they interpret (WRHA, 2021).

PIs can also help craft privacy notifications and user consent flows in the UX, ensuring users are aware that their speech is being recorded or processed. For instance, a professional interpreter would know that in a medical scenario, patients should be informed if an AI is interpreting and possibly storing their medical information.

This can spur features like on-device interpretation (to keep data off external servers) and auto-deletion of audio after processing, areas where professional interpreter input has been valuable. Moreover, PIs can spot subtle privacy issues: perhaps the system announces the interpretation out loud when it should have sent it via a private earpiece, or it retains names in its phrase history that could be seen by others.

A case in point: emergency services initially hesitated to use S2ST systems due to data confidentiality issues when handling refugees' personal information. By designing systems like BabelDr (a speech translator with predefined phrases that operates securely on-site), developers addressed those concerns, likely drawing on professional interpreter and medical ethicist consultations to ensure no sensitive data was improperly shared (Spechbach et al., 2019).

8.4. Real-world deployment

8.4.1. *Fine-tuning models for domain-specific applications*

When deploying speech translation in specialized fields (medical, legal, technical, etc.), generic models often falter on terminology and protocol. In such a scenario, domain-expert PIs become crucial collaborators.

These are PIs who work in the field (medical PIs, court PIs, etc.) and understand the jargon and the high-stakes nature of communication. Their role in development is to fine-tune or customize the S2ST system for the domain.

This can involve curating a domain-specific glossary for the translation engine, providing example translations of technical terms, or even recording sample utterances (for acoustic tuning) with the right parlance. For instance, in healthcare, a professional interpreter knows that a term like “heart attack” might need to be translated with a specific phrase in another language to be understood by lay patients, and that certain questions have to be asked in a particular way for clarity.

Projects like the aforementioned BabelDr illustrate this. It is a speech-enabled fixed-phrase translator designed for emergency medicine, developed through a collaboration between a hospital and a university institute of translation (Spechbach et al., 2019).

PIs and translators helped compile the list of critical diagnostic questions in multiple languages, ensuring both medical accuracy and patient-friendly wording. The result was a tool that doctors could use to successfully diagnose patients when PIs were unavailable.

Similar initiatives have emerged in law enforcement and disaster response, where PIs collaborate with engineers to develop domain-specific interpretation (Karaman, 2024). These include practical phrases such as requesting identification or delivering clear instructions during emergencies.

By fine-tuning models with domain knowledge, often through human-in-the-loop training (e.g. a professional interpreter correcting the system’s output on real domain data), the S2ST system potentially becomes viable for real-world tasks. This targeted improvement has been shown to increase accuracy dramatically in those scenarios (Yang et al., 2023).

Moreover, PIs can define what constitutes a critical error in their domain, for a medical translator, confusing “allergy” with “cold” is far worse than a small grammatical mistake. This prioritization can help developers optimize the model for the errors that matter most.

8.4.2. Testing systems in live interpretation scenarios to improve robustness

Lab tests and benchmarks only go so far – the true test of a speech translator is a live multilingual conversation. PIs play a key role in such trials, both as participants and evaluators.

One effective approach involves shadowing, where a professional interpreter listens to a live event translated by the system and simultaneously provides their own translation, either silently or to a small control group. This method allows for real-time comparison and correction, facilitating the refinement of the S2ST system’s outputs through immediate human feedback.

By comparing the human professional interpreter’s output with the system’s output in real time, developers can pinpoint where the system struggles. The professional interpreter can then feedback specific issues (e.g. “the system did not handle the speaker’s joke correctly” or “it dropped the second part of the compound sentence when the speaker kept talking without pause”).

This approach served as a rigorous “stress test” for early S2ST systems like Skype Translator, which underwent extensive user trials involving bilingual speakers and PIs who monitored sessions to identify and correct misinterpretations. These trials provided critical feedback, enabling developers to refine the system’s accuracy and reliability in real-world multilingual communication scenarios (Lewis, 2015).

Another strategy is pilot deployment at events with backup PIs on call. For example, at some international conferences and webinars, organizers have tested AI speech translation to provide captions or audio in multiple languages.

PIs might be on standby to step in if the AI fails or to correct major errors on the fly. These pilots provide valuable data: if certain phrases consistently trip up the system in a live setting, they can be added to training or trigger a fallback.

A case study comes from the field of remote meetings: Interprefy reported supporting over 50,000 multilingual meetings including high-profile events like press conferences and even an astronaut interview. In 2023, they introduced an AI translator (Aivia) into their platform, and while it was a technological first, they acknowledged that AI and humans will play complementary roles – with humans overseeing nuance and sensitive content (Macauley, 2023).

Early uses of Aivia in live events serve to refine it; PIs review the AI's performance and help adjust it for the next event. This iterative testing under real conditions leads to more robust systems – capable of handling varied accents, audio conditions, and interactive dialogues.

Incorporating PIs into the feedback loop enables S2ST systems to learn from real-world errors, progressively narrowing the performance gap between automated translations and those produced by skilled human PIs. This human-in-the-loop approach not only enhances the system's accuracy but also ensures adaptability to the nuanced demands of multilingual communication in dynamic environments.

8.4.3. Addressing latency and real-time usability challenges

One of the toughest challenges in S2ST is achieving low delay while maintaining translation quality (Dugan et al., 2023). PIs are masters of this balancing act, especially simultaneous PIs who translate as the speaker is talking (Amos et al., 2022).

PIs develop tactics over decades to minimize lag: for example, splitting long sentences into shorter units, anticipation of where a sentence is going, or subtle reordering of phrases to start speaking the translation sooner (Smith, 2023; Kikтова et al., 2020). These human strategies offer a roadmap for system developers.

In fact, research comparing human simultaneous interpreting with machine output found that PIs regularly apply several effective tactics to reduce translation delay, including

sentence segmentation and clause restructuring. Engineers are now encoding similar strategies into algorithms (He et al., 2016).

One recent study explicitly drew inspiration from how PIs alternate listening and speaking – the so-called read/write policy – to create an AI that decides when to wait for more words and when to begin translating in order to sound as fluent as possible with minimal delay (Cheng et al., 2024). The result was significantly improved performance aligned with human preferences for low latency.

PIs can also directly advise on acceptable delay for different scenarios. For instance, in casual conversation delay is better tolerated than in fast-paced emergency communications. By understanding the context, PIs can guide developers on setting delay thresholds or offering modes like “fast (lower accuracy)” vs “full accuracy (slower)” to users.

PIs can also help design incremental output strategies. For example, they know it is often better to output a partial interpretation rather than wait to hear the entire description before starting to interpret, because listeners prefer getting information continuously. Implementing this requires careful segmentation which professional interpreter insight can inform.

This kind of collaboration has led to systems that approach simultaneous translation in a more human-like way. For example, Meta’s recent simultaneous translation research introduced a technique to drop or rephrase unimportant filler words, very much like a professional interpreter would, to save time and focus on key content (Barrault et al., 2023).

All these advancements can reduce the awkward silence that plagued earlier S2ST systems. By analysing and emulating professional interpreter strategies, S2ST developers can drastically improve real-time usability.

9. Results and Discussion

ChatGPT-o1 argues for a paradigm shift in S2ST design: moving from a system that simply attempts to translate correctly to one that actively manages communication breakdowns by mimicking PIs. The core thesis, that error handling should be conversational rather than static, seems robust. However, applying human strategies to S2ST interfaces introduces specific friction points regarding cognitive load, latency, and context awareness.

In this chapter we propose a critical discussion of the arguments previously presented by ChatGPT-o1, categorized by the opportunities they reveal and the implementation challenges they may hide.

9.1. Discussing Design and User Experience (UX)

ChatGPT-o1 advocates for a “human-in-the-loop” design philosophy for S2ST systems. It argues that the tacit knowledge of PIs, specifically regarding error management, inclusivity, and conversation flow, should be codified into S2ST architecture.

9.1.1. The shift from error reporting to “grounding”

ChatGPT-o1 argues that S2ST systems act too passively when errors occur (defaulting to generic messages) and should instead adopt active “grounding” strategies used by humans (e.g., partial paraphrasing or confirmation requests). The core argument is that humans rarely ask for verbatim repetition; they repair conversations contextually. Machines should do the same.

While the proposal to use “gentle clarification prompts” is sound, it introduces a new layer of complexity: meta-communication (talking about the conversation rather than having it) (Hoppenbrouwers & Weigand, 2000). A PI knows exactly when to interrupt to clarify a number without derailing the speaker’s train of thought (Vogler; Stewart; Neubig, 2019). A S2ST system that interrupts to ask, “Did you mean 50 or 15?”, can shatter the conversational flow if the timing is off by even a few milliseconds (Tissi, 2000).

The suggestion to use voice commands (like “scratch that”) enhances control but increases cognitive load (Seeber, 2013). The user must now act as both a participant in the conversation and an editor of the translation. This moves the S2ST system away from being an invisible medium toward being a tool that requires active management.

9.1.2. Accessibility and the "one size fits all" fallacy

ChatGPT-o1 identifies that pure voice interfaces fail specific demographics (e.g., heavy accents, speech impediments, or low tech-literacy) and suggests multimodal inputs (typing, icons) based on PI experiences. The primary value proposition of S2ST is often “hands-free” or “eyes-free” communication (Bach et al., 2007). Introducing typing or selecting illustrations (as suggested in section 8.1.1) solves accessibility issues but transforms the system into a standard text-translation app with voice features. The critical point is: is the goal to replace the interpreter (not that we want that) or to provide a communication aid?

Moreover, ChatGPT-o1 cites the need to include minority languages and dialects to avoid marginalization. While ethically imperative (Branchadell, 2012), this faces a significant technical hurdle: data scarcity (Dinh, Liu, Niehues, 2022). S2ST models require massive datasets (Duquenne et al., 2023).

Major languages work well because the data exists. Implementing the PI standard for regional dialects requires not just design intent (Ganeshan; Garrett; Finger, 1994), but a fundamental shift in how training data is collected and curated.

9.1.3. Interface Design and Human Mimicry

ChatGPT-o1 suggests using PIs in “Wizard-of-Oz” experiments to design cues for turn-taking and confidence display, aiming for a system that “behaves like a considerate professional interpreter.” There is, however, a risk in making a machine mimic human back-channeling (saying “mm-hmm”) (Placani, 2024). If a S2ST system uses human-like conversational fillers, users may overestimate its intelligence or comprehension capabilities, leading to trust violations when the system eventually fails (Abercrombie et al., 2023).

The suggestion to show confidence indicators is technically feasible but UX heavy. In a fluid conversation, does a user have time to check a percentage bar? PIs filter this uncertainty

internally and only flag it when necessary (Bartłomiejczyk, 2006). Offloading this verification to the users burdens them with decision-making that the PI usually handles.

Table 17 summarizes the tensions between the ideal and the technical reality of S2ST systems:

Table 17 - Tensions between the ideal and the technical reality of S2ST systems

Feature Proposed	PI Behavior (Ideal)	S2ST Challenge (Reality)
Repair Strategy	Contextual rephrasing / asking pointed questions.	Risk of hallucination (inventing context) or disruptive timing.
Inclusivity	Adapting to any accent/dialect instantly.	Profound lack of training data for non-standard dialects.
Turn-Taking	Intuitive management of silence and interruptions.	Latency issues make real-time turn-taking feel "jerky" or rude.
Verification	"Scratch that" / Proactive repetition.	Increases cognitive load; users must "edit" the AI live.

Source: the author.

ChatGPT-o1 presents a compelling roadmap for moving S2ST from translation tools to conversational partners. However, the critical gap lies in contextual awareness (Setton, 2006). A PI understands why a conversation is happening (e.g., a high-stakes medical diagnosis vs. a casual greeting) and adjusts their error handling accordingly (Bartłomiejczyk, 2006). Until S2ST systems possess this situational awareness, implementing human-like repair strategies may result in an interface that feels intrusive rather than helpful.

9.2. Discussing linguistic and cultural adaptation

ChatGPT-o1 advances the argument that S2ST must evolve from a dictionary-based utility into a culturally intelligent agent. It posits that the tacit knowledge of PIs, specifically in pragmatics, cultural mediation, and prosody, is the missing link for next-generation S2ST.

9.2.1. *The pragmatic gap*

According to ChatGPT-o1, machines often fail at discourse coherence, maintaining the thread of a conversation (e.g., remembering that “Dr. Smith” is a “she” ten sentences later) (Xiong et al., 2019). PIs excel here because they build a mental model of the people, not just the sentences (Gao, 2011).

While ChatGPT-o1 suggests PIs can “review transcripts” to fix this, the technical bottleneck is the context window. Most current S2ST systems process audio in short chunks (sentence-by-sentence) to minimize latency (Polák & Bojar, 2024). To achieve the discourse-level consistency, S2ST must maintain a memory bank of the entire conversation history in real-time, which is computationally expensive and increases lag (Hussein et al., 2024).

ChatGPT-o1 implies the system should inject politeness markers. This raises an ethical question: should an application edit the speaker? If a speaker is being intentionally rude or abrupt, a PI might soften it to keep the peace (Felberg, 2016). If a S2ST does this automatically, it risks altering the speaker’s true intent, potentially masking conflict that needs to be communicated (e.g., in a legal or negotiation setting).

9.2.2. *Cultural mediation vs. hard-coding culture*

The argument that PIs can act as cultural liaisons to prevent errors (like the “good morning” vs. “attack them” incident) is compelling but difficult to scale. ChatGPT-o1 proposes curated cultural databases. However, culture is fluid (Zhou; Bamman; Bleaman, 2025). Slang, taboos, and political sensitivities shift rapidly (Keidar et al., 2022). A static database is likely to become outdated quickly (Chaieb; Hnich; Mrad, 2022).

The suggestion to have locale settings (e.g., Mexican vs. Peninsular Spanish) is standard in text localization but exponentially harder in S2ST (Zampieri; Nakov; Scherrer, 2020). It requires the system to not just swap words (vocabulary) but to swap concepts.

Automatically substituting a culturally equivalent phrase sounds efficient but risks erasing cultural specificity (Kabir et al., 2025). If a S2ST system translates a Japanese proverb about cherry blossoms into an American idiom about making hay, it erases the cultural flavor of the original speaker. PIs navigate this balance intuitively (Hajiyeva, 2025); algorithms may struggle to know when to domesticate a concept and when to leave it foreign for effect (Yao et al., 2024).

9.2.3. Prosody and emotion

ChatGPT-o1 advocates for preserving “how it is said” (prosody/emotion) alongside “what is said,” citing models like Meta’s Direct S2ST. Traditional S2ST systems are cascaded (ASR → MT → TTS). In this process, the MT (text) bottleneck strips away all emotion, tone, and irony (Sarim et al., 2025). ChatGPT-o1, however, identifies that direct S2ST (which maps audio to audio without a text middleman) is the solution.

ChatGPT-o1 assumes that preserving the original speaker’s pitch/emotion is always desirable. This is a fallacy. For example, a neutral/polite tone in US English often involves higher pitch and variation. In some Nordic or East Asian contexts, that same pitch variation might be perceived as childish or insincere (Brown; Oh; Idemaru, 2025).

A PI does not just copy the emotion; they transcreate it (Hsieh & Nicodemus, 2015). If the S2ST simply clones the acoustic features (voice cloning), it might accidentally convey the wrong emotion in the target culture.

Table 18 presents a summary of the mediator dilemma presented by ChatGPT-o1.

Table 18 - The "Mediator" Dilemma.

Feature	PI (Mediator)	S2ST System (Translator)	The Critical Tension
---------	---------------	--------------------------	----------------------

Politeness	Adds/removes honorifics based on social hierarchy.	Usually defaults to a “safe” standard (neutral).	If S2ST adds politeness the speaker didn’t intend, is it lying?
Culture	Explains or adapts idioms (e.g., “It’s raining cats and dogs” → “It’s pouring”).	Often literal or fails to recognize the idiom.	Cultural databases are hard to maintain and can feel stereotypical.
Emotion	Modulates voice to fit <i>target</i> culture's norms for that emotion.	Voice cloning copies source acoustics.	Direct cloning can sound “foreign” or “manic” in the target language.

Source: the author?

ChatGPT-o1 presents a strong vision for human-centric S2ST, but it arguably underestimates the complexity of cultural transcreation. A PI does not just translate languages; they translate civilizations (Pistillo, 2003).

While S2ST can certainly improve by incorporating prosody tokens and larger context windows, the ultimate goal of behaving like a PI requires the system to have world models, a fundamental understanding of social hierarchies, history, and human psychology, which remains one of the hardest problems in Computing (Gupta & Pruthi, 2025).

9.3. Discussing Ethical Considerations

ChatGPT-o1 positions PIs as the ethical conscience of S2ST development, arguing that their adherence to codes of ethics (confidentiality, impartiality, accuracy) should be encoded into S2ST behavior. Below is a critical discussion of these arguments, contrasting the ethical imperative with the computational and economic realities of deploying these systems.

9.3.1. Bias mitigation breaks conversation

ChatGPT-o1 argues that PIs can help flag distortions (like gender bias) and suggests mitigation strategies like presenting both gender options or randomly varying pronouns. The suggestion to present both gender options (e.g., The doctor [he/she]...) is a text-first solution. In S2ST, this is catastrophic for user experience. An audio output that says “The doctor—he or she—said...” disrupts the natural flow and doubles the length of the utterance.

The proposal to randomly vary gender to prevent stereotypes is technically feasible but pragmatically dangerous (Gonen & Goldberg, 2019). In a sustained conversation about a

specific individual, consistency is key (Welleck et al., 2019). If a S2ST system refers to a female doctor as “she” in sentence one, and then randomly switches to “he” in sentence three to balance the dataset, it creates confusion, not equity.

PIs flagging errors is a post-hoc solution. The bias exists in the word embeddings (the mathematical representation of words) derived from massive, biased internet scrapes (Bolukbasi et al., 2016). Auditing the output is comparable to filtering a contaminated river with a household water pitcher: meaningful improvement demands intervention at the source, the training data itself, a task that remains vast, complex, and largely unsolved in data science (Mehrabi et al., 2021).

9.3.2. The economics of universal fairness

ChatGPT-o1 notes that fairness means equitable quality across languages (e.g., Swahili vs. Spanish) and suggests PIs can push for gaps to be addressed. The disparity in quality between English-Spanish and Swahili-French is not just a lack of care; it is a function of data scarcity (Haddow et al., 2022).

S2ST models require petabytes of text/audio pairs to function (Rashidi & Sameti, 2025). For low-resource languages, this data simply does not exist in digital form (McGiff & Nikolov, 2025). PIs can advocate for better coverage, but without a massive, likely unprofitable, investment in data collection, the gap will remain.

Chapter 8 suggests ensuring the system does not consistently lag more for one language. This ignores linguistic typology. S2ST must lag more for some language pairs. For example, translating English (SVO) to Japanese (SOV). The verb comes at the end of the sentence in Japanese (Doi et al., 2024).

The system must wait until the English speaker finishes the sentence to know the verb, inevitably causing higher latency than translating English to French (Grissom et al., 2014). In such a scenario, fairness clashes with the laws of syntax.

9.3.3. Privacy: The edge vs. cloud trade-off

ChatGPT-o1 emphasizes the PI’s role in confidentiality and cites BabelDr (on-device, predefined phrases) as a success story for privacy. There is currently a direct trade-off between

privacy and intelligence (Avent et al., 2019). Cloud Models: high accuracy, nuance, and vocabulary (high privacy risk). On-Device (Edge) Models: secure, local processing. (lower accuracy, limited vocabulary, battery drain) (Pamadi & Singh, 2025).

ChatGPT-o1 champions BabelDr, but fails to mention that it relies on predefined phrases (Bouillon & Spechbach, 2016). It is not a generative translator; it is a fancy phrasebook. For true S2ST (generative, free-flowing conversation), we are still largely dependent on the cloud, making the confidentiality argument much harder to solve than ChatGPT-o1 implies.

PIs acting as guardians of consent (informing patients) is ideal, but in a consumer app, this usually degrades into a Terms of Service checkbox that no one reads (Steinfeld, 2016). The design challenge is how to make data privacy intuitive (e.g., a visual indicator of cloud processing vs. local mode) rather than just a legal disclaimer.

Table 19 shows the ethical ideals compared to technical constraints.

Table 19 - Ethical Ideals vs. Technical Constraints

Ethical Proposal	Technical Challenge
Present both gender options	Works in text; ruins the flow in spoken audio.
Randomly vary gender	Destroys coherence in longer conversations about specific people.
Equal latency for all languages	Impossible due to syntactic differences (e.g., verb-final vs. verb-initial languages).
Total privacy (on-device)	Currently results in significantly lower translation quality compared to cloud models.

Source: the author.

ChatGPT-o1 successfully identifies the moral requirements for S2ST, but glosses over the architectural hurdles. PIs are indispensable for defining what “good” looks like, but they cannot solve how to achieve it alone. Realizing this vision requires a shift in S2ST development from Big Data (scraping everything) to Smart Data (curated, balanced, and verified datasets), a shift that is currently slow and expensive (Iafrate, 2014).

9.4. Discussing real-world deployment

ChatGPT-o1 argues that the last mile of S2ST, bridging the gap between technically correct and professionally usable, requires the active collaboration of domain experts. It suggests that PIs should evolve from competitors into trainers, evaluators, and architectural blueprints for S2ST.

9.4.1. *The high-stakes paradox: fixed phrases vs. generative S2ST*

As pointed out earlier in this chapter, ChatGPT-o1 highlights the success of BabelDr as proof of PI collaboration. However, this example inadvertently exposes a major limitation of current generative S2ST.

BabelDr is effective precisely because it removes the generative aspect of AI-based S2ST. It relies on pre-vetted, static sentences (Spechbach et al., 2019). This suggests that for high-stakes domains (like emergency medicine), the industry still trusts a fancy phrasebook more than a neural network. The challenge remains: How do we bring the safety of BabelDr to open-ended S2ST?

The most valuable insight here is the PI’s ability to define critical errors. Standard metrics (like BLEU or WER) treat all word errors equally (Mathur; Baldwin; Cohn, 2020). To a S2ST, swapping “hypotension” (low blood pressure) for “hypertension” (high blood pressure) is a minor one-word error. To a doctor, it is fatal. PIs can help create semantic loss functions that penalize these specific medical/legal errors heavily during training (Birch et al., 2016).

9.4.2. *Live testing and the economics of redundancy*

ChatGPT-o1 proposes shadowing (PIs translating alongside S2ST to check it) and standby models (PIs stepping in if S2ST fails) as robustness measures. While standby PIs ensure safety, they negate the primary economic driver of any technology: cost reduction. If you must pay a professional to be on standby for every S2ST session, the cost savings vanish.

This model is likely a transitional phase, useful for gathering data to train the model, but economically unsustainable as a long-term service model.

Shadowing a S2ST to catch errors is cognitively different from interpreting. It requires the PI to function as an editor rather than a creator. Research suggests that post-editing (fixing bad output) can sometimes be more mentally taxing than translating from scratch (Koponen et al., 2012), potentially leading to alert fatigue where the PI misses errors because they are lulled into a false sense of security by the system's generally good performance.

9.4.3. Latency and the prediction trap

ChatGPT-o1 discusses minimizing delay by mimicking PI strategies like segmentation (chopping sentences) and anticipation (guessing the end of the sentence). PIs can anticipate (translate before hearing the verb) because they have a world model, they know that “the patient was treated for...” is likely followed by a condition (Amos; Seeber; Pickering, 2022). S2ST anticipates based on statistical probability (Ouyang et al., 2025).

However, if a system anticipates incorrectly in a casual chat, it's awkward. If it anticipates incorrectly in a legal deposition or medical triage, it is negligence. Hallucination is the dark side of anticipation (Kalai & Vempala, 2024).

ChatGPT-o1 praises dropping filler words (e.g., "um," "uh," "basically"). However, in legal contexts (like witness testimony), filler words often indicate hesitation or deception (Arciuli; Mallard; Villar, 2010). A PI knows when “um” matters; a S2ST might clean it up to make the transcript look pretty, inadvertently stripping the testimony of its credibility cues.

Table 20 summarizes the implications of PI as the gold standard.

Table 20 - The PI as the gold standard.

PI Strategy	Goal	Implementation	Risk
Anticipation	Reduce lag (speak before speaker finishes).	Predictive Decoding (guessing the next word).	Hallucination: AI may guess wrong, creating confident but false translations.

Simplification	Ensure clarity for lay audiences (e.g., patient).	Text Simplification algorithms.	Over-correction: AI might simplify technical terms that <i>need</i> to be precise (e.g., “myocardial infarction” → “heart hurt”).
Shadowing	Real-time quality control.	Online Learning (user feedback loops).	Economic Redundancy: Paying a human to watch a machine destroys the cost benefit.

Source: the author.

ChatGPT-o1 successfully argues that S2ST is no longer a raw engineering problem; it is a design and workflow problem. The brute force era of feeding generic data into massive models is yielding diminishing returns (Luo, 2024). The next leap in quality will come from Specialized Fine-Tuning, teaching the model not just the language, but PIs’ behavior (when to wait, when to clarify, and when to simplify).

However, ChatGPT-o1 glosses over the fundamental difference between human intuition and machine statistics. A human anticipates based on understanding; a machine anticipates based on probability (Keller & Masic-Flogel, 2018). Until S2ST systems can reason about the consequences of an error, mimicking human speed strategies (like anticipation) remains a dangerous game in high-stakes fields.

10. Conclusion

This dissertation set out to examine the role of PIs in S2ST beyond their conventional positioning as providers of training data and evaluators of system output. Drawing on two extensive, domain-specific corpora and on AI-assisted semantic analysis under strict human supervision, the study identified both the core domains of interpreter competence and the main stages in the development of S2ST systems.

By mapping these two dimensions, the research demonstrates that interpreter competence constitutes a form of epistemic and practical expertise that can inform multiple phases of S2ST design, implementation, and evaluation. In line with Pöchhacker's (2022) longstanding view of Interpreting as an embodied, situated, and socially embedded practice, the findings reinforce the argument that Interpreting cannot be reduced to linguistic transfer alone.

We critically examine ChatGPT-o1's proposal to redesign S2ST systems by modeling them after PIs. While the vision is compelling, reframing S2ST from a passive translation tool into an active conversational mediator, the analysis shows that many PI strategies clash with current technical, cognitive, and economic constraints.

As Pöchhacker (2022) has consistently emphasized, Interpreting is a form of meaning-making shaped by interactional dynamics, cultural context, ethical responsibility, and professional judgment. These dimensions, which remain largely inaccessible to purely data-driven S2ST models, position PIs not merely as benchmarks against which machine performance is measured, but as potential co-designers of communicative technologies.

ChatGPT-o1 correctly identifies PIs as the benchmark for quality, ethics, and usability in S2ST. However, it underestimates the depth of contextual, cultural, and moral reasoning required to replicate interpreter behavior. Progress will depend less on scaling data and more on carefully curated, domain-specific training and design choices that acknowledge, rather than obscure, the limits of current S2ST systems.

This study therefore extends Pöchhacker's (2024) critique of the limited interaction between Interpreting scholars and computational researchers by offering a concrete framework through which such interaction can be operationalized. The analysis shows that PI competence is relevant not only at the stages of corpus compilation and output evaluation, but also in system specification, error typology design, user interaction modeling, prosodic and pragmatic tuning, and ethical risk assessment.

In this sense, PIs emerge as domain experts in communicative risk management rather than as peripheral contributors to technological workflows. This reconceptualization aligns with Pöchhacker's (2024) caution against viewing S2ST as a functional equivalent of human interpreting, while simultaneously acknowledging the instrumental value of S2ST as a communicative aid in low- to medium-stakes contexts.

More broadly, this dissertation argues for a shift from a substitution-oriented paradigm to a complementarity-oriented model of human-machine interaction in speech translation. Such a model recognizes that S2ST systems gain legitimacy, robustness, and social acceptability

when informed by professional interpretive knowledge, just as the interpreting profession stands to benefit from active involvement in the shaping of emerging technologies. By articulating where, how, and why interpreter competence can contribute across the S2ST lifecycle, this study responds directly to Pöchlacker's (2024) call for deeper conceptual exchange between Interpreting Studies and computational translation research.

Finally, the framework proposed here is intended not as a prescriptive blueprint, but as a heuristic for future interdisciplinary research. It invites further empirical validation, refinement in specific interpreting settings, and closer collaboration between PIs, linguists, and technologists. In doing so, it contributes to a more human-centered vision of S2ST, one that treats technology not as a replacement for PIs, but as a site where their expertise can be meaningfully and productively inscribed.

11. Limitations

This dissertation presents limitations that should be acknowledged when interpreting its findings and implications. First, the study is based exclusively on secondary data drawn from academic publications indexed in Google Scholar and retrieved through Publish or Perish.

Although careful filtering and manual validation were carried out, the reliance on open-access literature may have led to the exclusion of relevant studies published behind paywalls or indexed in other databases. As a result, the compiled corpora may not fully represent the entire body of research on Interpreting or S2ST.

Second, the analytical core of the study depends on AI-assisted semantic and contextual analysis performed by ChatGPT-o1. While measures were taken to mitigate hallucinations, such as restricting the model's input to domain-specific corpora and manually verifying outputs, the possibility of omission, misinterpretation, or overgeneralization cannot be entirely ruled out. The findings should therefore be understood as the outcome of a human-AI collaborative process rather than a fully autonomous analysis.

Third, the study adopts a conceptual and exploratory approach. It does not include empirical validation through experiments, system prototyping, user studies, or direct observation of PIs working within S2ST development environments. Consequently, the proposed mappings between PI competence and S2ST development stages remain theoretical and illustrative, rather than empirically tested.

Fourth, the notion of “competence” is treated as an analytical construct synthesized from the literature. Although this choice is theoretically justified, it may oversimplify the diversity of interpreter profiles, working conditions, language pairs, and institutional settings. Competence is context-dependent, and the framework proposed here may not apply uniformly across all interpreting modes or cultural contexts.

Finally, the rapid evolution of artificial intelligence and S2ST technologies constitutes an inherent limitation. Some technical descriptions and system development stages discussed in this dissertation may become outdated as new architectures, evaluation methods, and deployment paradigms emerge. The study should therefore be read as a snapshot of a moving field rather than a definitive account.

Despite these limitations, the dissertation offers a structured framework and a novel interdisciplinary perspective that can inform future empirical research, system design, and collaboration between PIs and S2ST developers.

REFERENCES

- Abercrombie, G., Curry, A. C., Dinkar, T., Rieser, V., & Talat, Z. (2023). Mirages: On anthropomorphism in dialogue systems. *arXiv preprint arXiv:2305.09800*. <https://arxiv.org/abs/2305.09800>
- Agranovich, A., Nachmani, E., Rybakov, O., Ding, Y., Jia, Y., Bar, N., Zen, H., & Ramanovich, M. T. (2024). *SimulTron: On-device simultaneous speech-to-speech translation* [Preprint]. arXiv. <https://arxiv.org/abs/2406.02133>.

AIIC Professional interpreters. (2021). *With a little help from my Friends: AI for professional interpreter training and support* [Video]. YouTube. <https://www.youtube.com/watch?v=lQO5VSWZInY>.

Alarcón, S. M. (2025, March 20). Why subject matter expertise makes all the difference. LinkedIn. <https://www.linkedin.com/pulse/why-subject-matter-expertise-makes-all-difference-sergio-m-alarcon-r4rkc>. Retrieved 28 March 2025.

Alastruey, B., Sant, A., Gállego, G. I., Dale, D., & Costa-jussà, M. R. (2023). Speechalign: a framework for speech translation alignment evaluation. *arXiv preprint arXiv:2309.11585*. <https://arxiv.org/abs/2309.11585>.

Alawi, N. M., & Abdulhaq, S. Y. (2017). *Machine Translation: The Cultural and Idiomatic Challenge*. <https://staff.najah.edu/en/publications/9709/>.

Almahasees, Z., Al-Natour, M., Mahmoud, S., & Aminzadeh, S. (2024). Bridging Communication Gaps in Crisis: A Case Study of Remote Interpreting in the Middle East During the COVID-19 Pandemic. *World Journal of English Language*, 14(2), 462. https://www.researchgate.net/profile/Zakaryia-Almahasees/publication/377976515_Bridging_Communication_Gaps_in_Crisis_A_Case_Study_of_Remote_Interpreting_in_the_Middle_East_During_the_COVID-19_Pandemic/links/65c10d2534bbff5ba7ef817a/Bridging-Communication-Gaps-in-Crisis-A-Case-Study-of-Remote-Interpreting-in-the-Middle-East-During-the-COVID-19-Pandemic.pdf.

Almeida, P. M. R. D., & Nordin, J. N. (2017). Interpretação forense: a experiência prática da Justiça Federal de Guarulhos e o treinamento de intérpretes. *Ajufe, São Paulo, ano, 30*, 481-520. <https://www.academia.edu/download/87449685/226967635.pdf>.

Al Sharou, K., & Specia, L. (2022, June). A taxonomy and study of critical errors in machine translation. In Proceedings of the 23rd annual conference of the European Association for Machine Translation (pp. 171-180). <https://aclanthology.org/2022.eamt-1.20/>

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., ... & Zhu, Z. (2016, June). Deep speech 2: End-to-end speech recognition in English and Mandarin. In International conference on machine learning (pp. 173-182). PMLR. <http://proceedings.mlr.press/v48/amodei16.html?ref=https://codemonkey.link>

Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition*, 220, 104987. <https://www.sciencedirect.com/science/article/pii/S0010027721004108>

Andrews, T. (2025, January 15). *What is a terminology management system?* Professional interpreters & Translators, Inc. <https://ititranslates.com/what-is-a-terminology-management-system/>. Retrieved 28 March 2025.

Angelelli, C. V. (2001). *Deconstructing the invisible professional interpreter: A critical study of the interpersonal role of the professional interpreter in a cross-cultural/linguistic communicative event*. Stanford University. <https://search.proquest.com/openview/96a425480939216d845883faaa7e5b6a/1?pq-origsite=gscholar&cbl=18750&diss=y>.

Angelelli, C. V. (2004). *Medical interpreting and cross-cultural communication*. Cambridge University Press. <https://books.google.com.br/books?hl=pt-BR&lr=&id=SzWDslOjwwkC&oi=fnd&pg=PP1&dq=Angelelli+2004&ots=T1GI6CWIQ&sig=OudTvngu4J8JmCGGzamGmCbqKlg>.

Angelelli, C. V. (2011, March). The interpersonal role of the professional interpreter in cross-cultural communication: A survey of conference, court and medical professional interpreters in the US, Canada and Mexico. In *The Critical Link 3: Professional interpreters in the Community. Selected papers from the Third International Conference on Interpreting in Legal, Health and Social Service Settings, Montréal, Quebec, Canada 22–26 May 2001* (pp. 15-26). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.46.06ang/html>.

Angermeyer, P. S. (2010). Professional interpreter-mediated interaction as bilingual speech: Bridging macro-and micro-sociolinguistics in codeswitching research. *International journal of bilingualism*, 14(4), 466-489. <https://journals.sagepub.com/doi/abs/10.1177/1367006910370914>

Araabi, A., & Monz, C. (2020). Optimizing transformer for low-resource neural machine translation. *arXiv preprint arXiv:2011.02266*. <https://arxiv.org/abs/2011.02266>.

Avent, B., González, J., Diethe, T., Paleyes, A., & Balle, B. (2019). Automatic discovery of privacy-utility pareto fronts. *arXiv preprint arXiv:1905.10862*. <https://arxiv.org/abs/1905.10862>

Ayan, I. (2020). Re-thinking neutrality through emotional labour: The (in) visible work of conference professional interpreters. *TTR*, 33(2), 125-146. <https://www.erudit.org/en/journals/ttr/2020-v33-n2-ttr06076/1077714ar/abstract/>.

Azeroual, O., Saake, G., & Schallehn, E. (2018). Analyzing data quality issues in research information systems via data profiling. *International Journal of Information Management*, 41, 50-56. <https://www.sciencedirect.com/science/article/pii/S0268401218300975>.

Aziz, A. F. A., Tiun, S., & Ruslan, N. (2023). End to end text to speech synthesis for malay language using tacotron and tacotron 2. *International Journal of Advanced Computer Science and Applications*, 14(6). https://www.researchgate.net/profile/Noraini-Ruslan-3/publication/372171671_End_to_End_Text_to_Speech_Synthesis_for_Malay_Language_using_Tacotron_and_Tacotron_2/links/64d06b9640a524707b9c22ad/End-to-End-Text-to-Speech-Synthesis-for-Malay-Language-using-Tacotron-and-Tacotron-2.pdf

Babcock, L., Capizzi, M., Arbula, S., & Vallesi, A. (2017). Short-term memory improvement after simultaneous interpretation training. *Journal of Cognitive Enhancement*, 1, 254-267. <https://link.springer.com/article/10.1007/s41465-017-0011-x>.

Bach, N., Eck, M., Charoenpornasawat, P., Köhler, T., Stüker, S., Nguyen, T., ... & Black, A. W. (2007). The CMU TransTac 2007 eyes-free two-way speech-to-speech translation system. In *Proceedings of the Fourth International Workshop on Spoken Language Translation*.

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33, 12449-12460. <https://proceedings.neurips.cc/paper/2020/hash/92d1e1eb1cd6f9fba3227870bb6d7f07-Abstract.html>

Bahar, P., Bieschke, T., Schlüter, R., & Ney, H. (2021, January). Tight integrated end-to-end training for cascaded speech translation. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 950-957). IEEE. <https://ieeexplore.ieee.org/abstract/document/9383462/>

Baigorri-Jalón, J. (1999). Conference interpreting: From modern times to space technology. *Interpreting*, 4(1), 29-40. <https://doi.org/10.1075/intp.4.1.05bai>.

Baigorri-Jalon, J. (2010, December). Wars, languages and the role (s) of professional interpreters. In *Les liaisons dangereuses: Langues, traduction, interprétation* (No. 24, pp. p-173). <https://confremo.hal.science/hal-00599599/>.

Baigorri-Jalón, J. (2014). From Paris to Nuremberg. <https://www.torrossa.com/gs/resourceProxy?an=5001144&publisher=FZ4850>.

Baigorri-Jalón, J. (2015). The history of the interpreting profession. In *The Routledge handbook of interpreting* (pp. 11-24). Routledge. <https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.4324/9781315745381-3&type=chapterpdf>.

Baixauli-Olmos, L. (2020). Ethics codes for professional interpreters and translators. In *The Routledge handbook of translation and ethics* (pp. 297-319). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003127970-23/ethics-codes-professional-interpreters-translators-llu%C3%ADs-baixauli-olmos>.

Baker, M. (2018). *Translation and conflict: A narrative account*. Routledge. <https://www.taylorfrancis.com/books/mono/10.4324/9780429438240/translation-conflict-mona-baker>.

Baker, M., & Diriker, E. (2019). Conference and simultaneous interpreting. In *Routledge encyclopedia of translation studies* (pp. 95-101). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315678627-21/conference-simultaneous-interpreting-mona-baker-ebru-diriker>.

Bangalore, S., Sridhar, V. K. R., Kolan, P., Golipour, L., & Jimenez, A. (2012, June). Real-time incremental speech-to-speech translation of dialogs. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 437-445). <https://aclanthology.org/N12-1048.pdf>.

Barcelo Jr, C. J. (1979). The Court Professional interpreters Act--A Step towards Equal Justice. *Agenda*, 9(3), 21. <https://eric.ed.gov/?id=EJ206916>.

Barrault, L., Chung, Y. A., Meglioli, M. C., Dale, D., Dong, N., Duquenne, P. A., ... & Wang, S. (2023). SeamlessM4T: Massively Multilingual & Multimodal Machine Translation. *arXiv preprint arXiv:2308.11596*. <https://doi.org/10.48550/arXiv.2308.11596>

Bartłomiejczyk, M., & Stachowiak-Szymczak, K. (2021, November). Modes of conference interpreting: Simultaneous and consecutive. In *The Routledge handbook of conference interpreting* (pp. 19-33). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429297878-4/modes-conference-interpreting-magdalena-bart%C5%82omiejczyk-katarzyna-stachowiak-szymczak>.

Bartłomiejczyk, M. (2006). Strategies of simultaneous interpreting and directionality. *Interpreting*, 8(2), 149-174. <https://www.jbe-platform.com/content/journals/10.1075/intp.8.2.03bar>

Bawden, R. (2018). *Going beyond the sentence: Contextual machine translation of dialogue* (Doctoral dissertation, Université Paris Saclay (COMUE)). <https://theses.hal.science/tel-02004683/>

Baxter, R. N. (2019). An overview of interpreting in Spanish: Past, present and future. *The Routledge Handbook of Spanish Translation Studies*, 196-210. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315520131-12/overview-interpreting-spanish-robert-neal-baxter>.

Behrend, T. S., Ravid, D. M., & Rudolph, C. W. (2024). Technology and the changing nature of work. *Journal of Vocational Behavior*, 154, 104028. <https://www.sciencedirect.com/science/article/pii/S0001879124000691>.

Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., Fissore, L., Laface, P., Mertins, A., Ris, C., Rose, R., Tyagi, V., & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, 49(10–11), 763–786. <https://doi.org/10.1016/j.specom.2007.07.010>

Berk-Seligson, S. (2017). *The bilingual courtroom: Court professional interpreters in the judicial process*. University of Chicago Press.

Biadsy, F., Weiss, R. J., Moreno, P. J., Kanevsky, D., & Jia, Y. (2019). Parrotron: An end-to-end speech-to-speech conversion model and its applications to hearing-impaired speech and speech separation. *arXiv preprint arXiv:1904.04169*. <https://arxiv.org/abs/1904.04169>

Biagini, M., Boyd, M., & Monacelli, C. (2017). *The Changing Role of the Professional interpreter*. Routledge. <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315621531&type=googlepdf>.

Birch, A., Abend, O., Bojar, O., & Haddow, B. (2016, November). HUME: Human UCCA-based evaluation of machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1264–1274). <https://aclanthology.org/D16-1134.pdf>

Bolden, G. B. (2000). Toward understanding practices of medical interpreting: Professional interpreters' involvement in history taking. *Discourse studies*, 2(4), 387–419. <https://journals.sagepub.com/doi/abs/10.1177/1461445600002004001>.

Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, 29. https://proceedings.neurips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html

Bostrom, K., & Durrett, G. (2020). Byte pair encoding is suboptimal for language model pretraining. *arXiv preprint arXiv:2004.03720*. <https://arxiv.org/abs/2004.03720>.

Bouillon, P., & Spechbach, H. (2016). BabelDr: a web platform for rapid construction of phrasebook-style medical speech translation applications. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation: Projects/Products*. <https://aclanthology.org/2016.eamt-2.5.pdf>

Bowker, L. (2021). Promoting linguistic diversity and inclusion. *The International Journal of Information, Diversity, & Inclusion*, 5(3), 127–151. <https://www.jstor.org/stable/48644449>

Bown, S. (2013, May). Autopoiesis: Scaffolding the reflective practitioner toward employability. *Conference of Professional interpreter Trainers*. <http://wlv.openrepository.com/handle/2436/622620>.

Braaten, Oddmund. (2023 April). Exclusive: Swiss startup unveils ‘world-first’ AI translation service. <https://thenextweb.com/news/interprefy-startup-unveils-worlds-first-advanced-automated-speech-translation-service-for-online-and-live-eventsinterprefy#:~:text=profession,humans%20can%20play%20complementary%20roles>

Branchadell, A. (2012). Minority languages and translation. In *Handbook of translation studies* (pp. 97–101). John Benjamins Publishing Company.

Braun, S. (2019). Technology and interpreting. In *The Routledge handbook of translation and technology* (pp. 271-288). Routledge.

Braun, S., Slater, C., Gittins, R., Ritsos, P. D., & Roberts, J. C. (2013). Interpreting in Virtual Reality: designing and developing a 3D virtual world to prepare professional interpreters and their clients for professional practice. *New prospects and perspectives for educating language mediators*, 5, 93-120. https://www.academia.edu/download/68617919/2013_Interpreting_in_Virtual_Reality_Des.pdf.

Brazier, C., & Rouas, J. L. (2024). Usefulness of Emotional Prosody in Neural Machine Translation. *arXiv preprint arXiv:2404.17968*. <https://arxiv.org/abs/2404.17968>

Bronkhorst, J. (2011). The invisible professional interpreter. *Asiatische Studien/Études Asiatiques*, 65(1), 35-43. https://serval.unil.ch/resource/serval:BIB_56D4EEA90FE4.P001/REF.pdf.

Brown, L. (2015). Honorifics and politeness. *The handbook of Korean linguistics*, 303-319. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118371008.ch17>.

Brown, L., Oh, G. E., & Idemaru, K. (2025). Prosodic features of polite speech: Evidence from Korean interactional data. *Pragmatics*, 35(3), 321-347. <https://www.jbe-platform.com/content/journals/10.1075/prag.23003.bro>

Bulyko, I., Kirchhoff, K., Ostendorf, M., & Goldberg, J. (2005). Error-correction detection and response generation in a spoken dialogue system. *Speech Communication*, 45(3), 271-288. <https://www.sciencedirect.com/science/article/pii/S0167639304001153>

Burns, J. (2010). Stress busters for professional interpreters (and everyone else). *The ATA Chronicle*, 39(6), 24-27. https://www.imiaweb.org/uploads/docs/interpreterstress_julie_burns.pdf.

Buzungu, H. F., & Hansen, J. P. (2020). Bridging Divides in the Interpreting Profession: Response to Gile and Napier (2020). *International Journal of Professional interpreter Education*, 12(2), 7. <https://open.clemson.edu/ijie/vol12/iss2/7/>.

Camacho-Collados, J., & Pilehvar, M. T. (2017). On the role of text preprocessing in neural network architectures: An evaluation study on text categorization and sentiment analysis. *arXiv preprint arXiv:1707.01780*. <https://arxiv.org/abs/1707.01780>.

Carballido, J., Olen, A., Lim, P., Davies, W. H., & Nordness, K. (2024). Peer support for medical professional interpreters: Is confidentiality a barrier? *International Journal of Professional interpreter Education*, 15(1), Article 12. <https://open.clemson.edu/ijie/vol15/iss1/12>.

Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting*, 16(2), 139-168. <https://www.jbe-platform.com/content/journals/10.1075/intp.16.2.01tim>.

Chaieb, S., Hnich, B., & Mrad, A. B. (2022). Data obsolescence detection in the light of newly acquired valid observations. *Applied Intelligence*, 52(14), 16532-16554. <https://link.springer.com/article/10.1007/s10489-022-03212-0>

Chan, C. C. H., Lin, Y. C., & Shih, P. C. (2023, December). Natural language processing for supply chain with chat GPT. In *2023 IEEE 5th International Conference on Architecture, Construction, Environment and Hydraulics (ICACEH)* (pp. 28-31). IEEE. <https://ieeexplore.ieee.org/abstract/document/10452639/>.

Chan, V. (2024). Impact of technology on interpreting practice: a review of studies on technology and interpreting practice from 2013 to 2024. *Interactive Technology and Smart Education*, (ahead-of-print). <https://www.emerald.com/insight/content/doi/10.1108/ITSE-02-2024-0042>.

Chandler, D. (1994). The transmission model of communication. URL: <http://www.aber.ac.uk/media/Documents/short/trans.html#Top> (дата звернення: 24.07. 2013). <https://www.aber.ac.uk/media/Documents/short/trans.html>.

Chatoui, H., & Ata, O. (2021, June). Automated Evaluation of the Virtual Assistant in Bleu and Rouge Scores. In *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/9461351/>

Chen, S. (2017). The construct of cognitive load in interpreting and its measurement. *Perspectives*, 25(4), 640-657. <https://www.tandfonline.com/doi/abs/10.1080/0907676X.2016.1278026>.

Chen, P., Guo, Z., Haddow, B., & Heafield, K. (2023). Iterative translation refinement with large language models. *arXiv preprint arXiv:2306.03856*. <https://arxiv.org/abs/2306.03856>

Cheng, S., Huang, Z., Ko, T., Li, H., Peng, N., Xu, L., & Zhang, Q. (2024). Towards Achieving Human Parity on End-to-end Simultaneous Speech Translation via LLM Agent. *arXiv preprint arXiv:2407.21646*. <https://arxiv.org/html/2407.21646v2#:~:text=Translation%20%28SiST%29,historical%20context%2C%20and%20retrieved%20information>

Chilingaryan, K., Gorbatenko, O., & Gorbatenko, R. (2016, October). Training court professional interpreters in Germany. In *Proceedings of ADVED 2016 2nd International Conference on Advances in Education and Social Sciences* (pp. 10-12). <https://www.academia.edu/download/76638399/230.pdf>.

Ciordia, L. S. (2017). A conceptual and contemporary approach to the evolution of impartiality in community interpreting. *The Journal of Specialised Translation*, 28, 273-292. <http://www.phoenix3.ir/admin/News/file/56.pdf>.

Cokely, D., & Marschark, M. (2005). Shifting positionality: A critical examination of the turning point in the relationship of professional interpreters and the deaf community. *Interpreting and Professional interpreter Education*, 1. <https://parhamti.com/wp-content/uploads/2023/07/Sign-Language-Interpreting-and-Interpreting-Education-%E2%80%93-Directions-for-Research-and-Practice.pdf#page=20>.

Colin, J., & Morris, R. (1996). *Professional interpreters and the legal process*. Waterside Press. https://books.google.com.br/books?hl=pt-BR&lr=&id=NB3bCgAAQBAJ&oi=fnd&pg=PP1&dq=officer+of+the+court+impartial+professional+interpreters&ots=Yl_uStB5c5&sig=xAs73Pj-PToVGJVYaQX_U13OrMA.

Corpas Pastor, G. (2021). *Interpreting and technology: Is the sky really the limit?* <https://wlv.openrepository.com/handle/2436/624500>.

Costa-Jussa, M. R., & Fonollosa, J. A. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*, 32(1), 3-10. <https://www.sciencedirect.com/science/article/pii/S0885230814001077>

Crockford, D. (2012). *Json. ECMA International*. <https://www.json.org/json.pdf>.

Dall, R., Yamagishi, J., & King, S. (2014). Rating naturalness in speech synthesis: The effect of style and expectation. In *Speech Prosody 2014*. <https://www.research.ed.ac.uk/en/publications/rating-naturalness-in-speech-synthesis-the-effect-of-style-and-ex>

Dang, H. T., & Phan, X. H. (2022, October). Non-standard vietnamese word detection and normalization for text-to-speech. In *2022 14th international conference on Knowledge and Systems Engineering (KSE)* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/9953791/>.

Davidson, B. (2001). Questions in cross-linguistic medical encounters: The role of the hospital professional interpreter. *Anthropological quarterly*, 74(4), 170-178. <https://muse.jhu.edu/pub/35/article/2106/summary>.

Defrancq, B., & Collard, C. (2020). Using data from simultaneous interpreting contrastive linguistics. In *New approaches to contrastive linguistics: empirical and methodological challenges* (Vol. 336, pp. 159-183). De Gruyter. <https://www.degruyter.com/document/doi/10.1515/9783110682588-006/pdf?licenseType=restricted>.

Delisle, J., Woodsworth, J., & Woodsworth, J. (Eds.). (2012). *Translators through history: revised edition*. John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.101.13ch9/html>.

Delongová, M. (2010). *English intonation patterns expressing politeness and their cross-language perception* (Doctoral dissertation, Ph. D. dissertation on the field of linguistics.—Olomouc). <https://theses.cz/id/8grh2i/64725-171595099.pdf>

Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Benefits*, 2(2), 2022. <https://pdfs.semanticscholar.org/2bf0/2911cbc3bc24e43a910477a90363225bed95.pdf>.

Deng, D., & Xue, N. (2017). Translation divergences in Chinese–English machine translation: An empirical investigation. *Computational Linguistics*, 43(3), 521-565. <https://direct.mit.edu/coli/article-abstract/43/3/521/1584>

Dessalles, J. L. (2007). *Why we talk: The evolutionary origins of language*. Oxford University Press. <https://global.oup.com/academic/product/why-we-talk-9780199276233?cc=br&lang=en&>.

Dhawan, S. (2022). Speech to speech translation: Challenges and future. *International Journal of Computer Applications Technology and Research*, 11(03), 36-55. <https://www.academia.edu/download/81705448/ijcatr11031001.pdf>.

Ding, W., Bhagoji, A. N., Zhao, B. Y., & Zheng, H. (2024, April). Towards Scalable and Robust Model Versioning. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)* (pp. 592-611). IEEE. <https://ieeexplore.ieee.org/abstract/document/10516632/>

Dinh, T. A., Liu, D., & Niehues, J. (2022, May). Tackling data scarcity in speech translation using zero-shot multilingual machine translation techniques. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6222-6226). IEEE. <https://ieeexplore.ieee.org/abstract/document/9746815/>

Diriker, E. D., & Baker, M. (1998). Conference Interpreting as Situated Actio. In *Routledge Encyclopedia of Translation Studies*. Routledge. <https://research.manchester.ac.uk/en/publications/conference-interpreting-as-situated-actio>.

Doi, K., Ko, Y., Makinae, M., Sudoh, K., & Nakamura, S. (2024, August). Word order in English-Japanese simultaneous interpretation: Analyses and evaluation using chunk-wise monotonic translation. In *Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024)* (pp. 254-264). <https://aclanthology.org/2024.iwslt-1.30/>

Dong, Y., & Li, P. (2020). Attentional control in interpreting: A model of language control and processing control. *Bilingualism: Language and Cognition*, 23(4), 716-728.

<https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article/attentional-control-in-interpreting-a-model-of-language-control-and-processing-control/137EB8E3503975D452283BA42D12E60F>

Donovan, C. (2017). The place of the professional interpreter and interpreting in an institutional setting. *Clina*, 3(2), 91-113. <https://pdfs.semanticscholar.org/1045/7246d2bf619ebd7b4d4adca1376f8cc1b524.pdf>.

Dose, S. (2020). Professional interpreters' strategies for dealing with different source speech delivery rates: Form-vs meaning-based approaches. [https://ujcontent.uj.ac.za/esploro/outputs/journalArticle/Professional interpreters-strategies-for-dealing-with-different/9913651407691](https://ujcontent.uj.ac.za/esploro/outputs/journalArticle/Professional+interpreters-strategies-for-dealing-with-different/9913651407691)

Dugan, L., Wadhawan, A., Spence, K., Callison-Burch, C., McGuire, M., & Zordan, V. (2023). Learning when to speak: Latency and quality trade-offs for simultaneous speech-to-speech translation with offline models. *arXiv preprint arXiv:2306.01201*. <https://arxiv.org/abs/2306.01201>

Duquenne, P. A., Gong, H., Dong, N., Du, J., Lee, A., Goswami, V., ... & Schwenk, H. (2023, July). Speechmatrix: A large-scale mined corpus of multilingual speech-to-speech translations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 16251-16269). <https://aclanthology.org/2023.acl-long.899/>

Duret, J., O'Brien, B., Estève, Y., & Parcollet, T. (2023, December). Enhancing expressivity transfer in textless speech-to-speech translation. In *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 1-8). IEEE.

EBSCO. (2024). Immigration reform in the United States. <https://www.ebsco.com/research-starters/law/immigration-reform-united-states?>

Edwards, R., Temple, B., & Alexander, C. (2005). Users' experiences of interpreters: The critical role of trust. *Interpreting*, 7(1), 77-95. <https://www.jbe-platform.com/content/journals/10.1075/intp.7.1.05edw>

Esqueda, M. D. (2025). Scientific articles on translation training published in Brazil: A bibliometric study of the last 41 years (1981-2022). *Alfa*. <https://doi.org/10.1590/1981-5794-e18887t>

European Union. (2010). *Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings*. EUR-Lex. <https://eur-lex.europa.eu/eli/dir/2010/64/oj/eng>.

Fantinuoli, C. (2018). Interpreting and technology: The upcoming technological turn. *Interpreting and technology*, 1-12.

Farrell, T. J. (1985). *Orality and Literacy: The Technologizing of the Word*. <https://www.jstor.org/stable/357987>.

Federico, M., Enyedi, R., Barra-Chicote, R., Giri, R., Isik, U., Krishnaswamy, A., & Sawaf, H. (2020). From speech-to-speech translation to automatic dubbing. *arXiv preprint arXiv:2001.06785*. <https://arxiv.org/abs/2001.06785>.

Felberg, T. R. (2016). Impoliteness—a challenge to interpreters' professionalism. *FLEKS-Scandinavian Journal of Intercultural Theory and Practice*, 3(1). <https://journals.oslomet.no/index.php/fleks/article/view/1683>

Fernández-Bravo, E. A. (2019). *Professional interpreter role (self-) perception: a model and an assessment tool*. <https://repositorio.comillas.edu/xmlui/handle/11531/37509>.

Fernández-Sánchez, M. M. (2019). Understanding interpreting and diplomacy: Reflections on the early Cold War (1945–1963). *The Palgrave Handbook of Languages and Conflict*, 395-418. https://link.springer.com/chapter/10.1007/978-3-030-04825-9_18.

Freitas, F. S. (2016). *O Estado da Arte da Interpretação Automática: do pós-guerra aos apps de Tradução automática de Fala*. [Final paper, Universidade Federal de Uberlândia].

Freitas, F. S. (2022). *Aplicativos Móveis de Interpretação Automática: expectativas e percepções dos usuários brasileiros*. [Master's thesis, Universidade Federal de Uberlândia]. Repositório da Universidade Federal de Uberlândia. <https://repositorio.ufu.br/handle/123456789/35038>

Freitas, F. de S.; Esqueda, M. D. (2017). Interpretação automática ou tradução automática de fala: conceitos, definições e arquitetura de software. *Tradterm*, v. 29, p. 104-145. <https://www.revistas.usp.br/tradterm/article/view/134416>.

Freitas, F. S. ; Esqueda, M. D. (2020). *Tradução e Interpretação Automáticas: Origens*. 1. ed. Curitiba: Editora CRV, v. 1. 104p.

Fuccellaro, M. (2024). *Concept Drift: detection, update and correction* (Doctoral dissertation, Université de Bordeaux). <https://theses.hal.science/tel-04954208/>

Fucci, D., Gaido, M., Papi, S., Cettolo, M., Negri, M., & Bentivogli, L. (2023). Integrating Language Models into Direct Speech Translation: An Inference-Time Solution to Control Gender Inflection. *arXiv preprint arXiv:2310.15752*. <https://arxiv.org/abs/2310.15752>

Fukuda, R., Sudoh, K., & Nakamura, S. (2023). Improving Speech Translation Accuracy and Time Efficiency with Fine-tuned wav2vec 2.0-based Speech Segmentation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 906-916.

Gaiba, F. (1998). *The origins of simultaneous interpretation: The Nuremberg Trial*. University of Ottawa Press.

Ganeshan, R., Garrett, J., & Finger, S. (1994). A framework for representing design intent. *Design Studies*, 15(1), 59-84. <https://www.sciencedirect.com/science/article/pii/0142694X94900396>

Gao, H. (2002). Language contact-Misunderstanding, confusion and conflicts. *Intercultural Communication Studies*, 11(3), 107-115. <https://www-s3-live.kent.edu/s3fs-root/s3fs-public/file/09-Hong-Gao.pdf>

Gao, Y., Cao, Z., Miao, Z., Yang, B., Liu, S., Zhang, M., & Su, J. (2024). Efficient k-Nearest-Neighbor Machine Translation with Dynamic Retrieval. *arXiv preprint arXiv:2406.06073*. <https://arxiv.org/abs/2406.06073>.

Gao, W. (2011). *Coherence in simultaneous interpreting: an idealized cognitive model perspective* (Doctoral dissertation, Heriot-Watt University). <http://www.ros.hw.ac.uk/handle/10399/2504>

Gastó Jiménez, M. (2018). The importance of prosodic elements in simultaneous interpreting. <https://repositori.upf.edu/items/3254a0e8-0e7b-4bcf-b1ba-74ebf2a8e5f5>

Geng, M., & Trotta, R. (2024). Is ChatGPT Transforming Academics' Writing Style? *arXiv preprint arXiv:2404.08627*. <https://arxiv.org/abs/2404.08627>.

Gieshoff, A. C. (2017). Audiovisual speech decreases the number of cognate translations in simultaneous interpreting. *Empirical modelling of translation and interpreting*, 7, 313. <https://library.oapen.org/bitstream/handle/20.500.12657/30853/641700.pdf?sequence=1#page=325>.

Gieshoff, A. C., & Albl-Mikasa, M. (2024). Interpreting accuracy revisited: a refined approach to interpreting performance analysis. *Perspectives*, 32(2), 210-228. <https://www.tandfonline.com/doi/abs/10.1080/0907676X.2022.2088296>.

Gile, D. (2006). Interpreting studies as an academic discipline: sociological and scientific aspects. *West and East: developments in translation studies.[In Chinese.] Baihua: Literature and Art Publishing House*, 283-301. https://www.academia.edu/download/87943330/2006_ISasacademdisc.pdf.

Gile, D. (2009). Interpreting studies: a critical view from within. *MonTI. Monografías de Traducción e Interpretación*, (1), 135-155. <https://www.redalyc.org/pdf/2651/265119728011.pdf>.

Gonen, H., & Goldberg, Y. (2019). Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *arXiv preprint arXiv:1903.03862*. <https://arxiv.org/abs/1903.03862>

Graham, C., & Roll, N. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *JASA Express Letters*, 4(2). <https://pubs.aip.org/asa/jel/article/4/2/025206/3267247>

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 6645-6649). IEEE. <https://ieeexplore.ieee.org/abstract/document/6638947/>

Gris, L. R. S., Fernandes, D., & de Oliveira, F. S. (2024, March). Automatic Speech-to-Speech Translation of Educational Videos Using SeamlessM4T and Its Use for Future VR Applications. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 163-166). IEEE. <https://ieeexplore.ieee.org/abstract/document/10536351/>

Grissom II, A., He, H., Boyd-Graber, J., Morgan, J., & Daumé III, H. (2014, October). Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)* (pp. 1342-1352). <https://aclanthology.org/D14-1140.pdf>

Gu, Y., Zhang, X., Xue, L., & Wu, Z. (2024, April). Multi-scale sub-band constant-q transform discriminator for high-fidelity vocoder. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 10616-10620). IEEE. <https://ieeexplore.ieee.org/abstract/document/10448436/>.

Guisse, K. (2020). Translating and Interpreting the Nuremberg Trials. *The National WWII Museum*, New Orleans. <https://www.nationalww2museum.org/war/articles/translating-and-interpreting-nuremberg-trials#:~:text=There%20are%20a%20few%20ways,simultaneous%2C%20but%20there%20is%20a>.

Gupta, R., Agarwal, U., & Helen Victoria, A. (2024, January). Multilingual Communication: NMT-Based On-Call Speech Translation for Indian Languages. In *International Conference on Multi-Strategy Learning Environment* (pp. 113-127). Singapore: Springer Nature Singapore. https://link.springer.com/chapter/10.1007/978-981-97-1488-9_9

Gupta, G., Rastegarpanah, B., Iyer, A., Rubin, J., & Kenthapadi, K. (2023). Measuring distributional shifts in text: the advantage of language model-based embeddings. *arXiv preprint arXiv:2312.02337*. <https://arxiv.org/abs/2312.02337>

- Gupta, T., & Pruthi, D. (2025). Beyond World Models: Rethinking Understanding in AI Models. *arXiv preprint arXiv:2511.12239*. <https://arxiv.org/abs/2511.12239>
- Haddow, B., Bawden, R., Miceli-Barone, A. V., Helcl, J., & Birch, A. (2022). Survey of low-resource machine translation. *Computational Linguistics*, 48(3), 673-732. <https://aclanthology.org/2022.cl-3.6/>
- Hajiyeva, B. (2025). Translating idioms and slang: Problems, strategies, and cultural implications. *Acta Globalis Humanitatis Et Linguarum*, 2(2), 284-293. <https://egarp.lt/index.php/aghel/article/view/201>
- Hale, S. (2014). Interpreting culture. Dealing with cross-cultural issues in court interpreting. *Perspectives*, 22(3), 321-331. <https://www.tandfonline.com/doi/abs/10.1080/0907676X.2013.827226>.
- Halley, M. (2020). Becoming allies: The transmogrification of the role of American Sign Language/English professional interpreters during the Deaf President Now protest. In *Disability Alliances and Allies* (Vol. 12, pp. 167-188). Emerald Publishing Limited. <https://www.emerald.com/insight/content/doi/10.1108/S1479-354720200000012012>.
- Haman, M., & Školník, M. (2024). Using ChatGPT to conduct a literature review. *Accountability in research*, 31(8), 1244-1246. <https://www.tandfonline.com/doi/full/10.1080/08989621.2023.2185514>.
- Hamed, I., Hussein, A., Chellah, O., Chowdhury, S., Mubarak, H., Sitaram, S., ... & Ali, A. (2023, January). Benchmarking evaluation metrics for code-switching automatic speech recognition. In *2022 IEEE Spoken Language Technology Workshop (SLT)* (pp. 999-1005). IEEE. <https://ieeexplore.ieee.org/abstract/document/10023181/>
- Harzing, A. W. (2010). *The publish or perish book*. Melbourne: Tarma Software Research Pty Limited. <http://www.harzing.info/download/popbook12.pdf>
- Hashana, A. J., Brundha, P., Ayoobkhan, M. U. A., & Fazila, S. (2023, April). Deep learning in ChatGPT-A survey. In *2023 7th international conference on trends in electronics and informatics (ICOEI)* (pp. 1001-1005). IEEE. <https://ieeexplore.ieee.org/abstract/document/10125852/>.
- Hayashi, T., Yamamoto, R., Inoue, K., Yoshimura, T., Watanabe, S., Toda, T., ... & Tan, X. (2020, May). ESPnet-TTS: Unified, reproducible, and integratable open source end-to-end text-to-speech toolkit. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 7654-7658). IEEE. <https://ieeexplore.ieee.org/abstract/document/9053512/>.
- He, H., Boyd-Graber, J., & Daumé III, H. (2016, June). Interpretese vs. translationese: The uniqueness of human strategies in simultaneous interpretation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 971-976). <https://www.semanticscholar.org/paper/Interpretese-vs.-Translationese%3A-The-Uniqueness-of-He-Boyd-Graber/37e57091f17cbc016e29cacbf90240b4d36db37#:~:text=Interpretation%20www,inclusing%20sentence%20segmentation%20and>
- Hinsvark, A., Delworth, N., Del Rio, M., McNamara, Q., Dong, J., Westerman, R., ... & Jette, M. (2021). Accented speech recognition: A survey. *arXiv preprint arXiv:2104.10747*. <https://arxiv.org/abs/2104.10747>
- Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. *Speech Communication*, 36(1-2), 31-43. <https://www.sciencedirect.com/science/article/pii/S0167639301000243>.
- Holloway, R. L. (1983). Human paleontological evidence relevant to language behavior. *Human neurobiology*, 2(3), 105-114. <https://www.columbia.edu/~rlh2/HneurobioLang1983.pdf>.

Hoppenbrouwers, S. J. B. A., & Weigand, H. (2000). Meta-communication in the language-action perspective. In *Proceedings of the 5th International Workshop on the Language-Action Perspective on Communication Modelling (LAP2000)* (pp. 131-149). RWTH Aachen.

Horiguchi, K. (1997). Towards translating spoken language pragmatics in an analogical framework. In *Spoken Language Translation*. <https://aclanthology.org/W97-0403.pdf#:~:text=speech%20translation%20is%20gaining%20importance,of%20the%20speaker%27s%20linguistic%20incompetence>

Horváth, I. (2022). AI in interpreting: Ethical considerations. *Across Languages and Cultures*, 23(1), 1-13. https://www.researchgate.net/publication/367709771_AI_in_interpreting_Ethical_considerations/fulltext/63da9597c97bd76a8253c8b9/AI-in-interpreting-Ethical-considerations.pdf.

House, B. (2017). Machine Listening: WaveNet, media materialism, and rhythm analysis. *A Peer-Reviewed Journal About*, 6(1), 16-24. <https://assets.pubpub.org/rhywrie2/31608661272658.pdf>

Hsieh, E. (2006). Conflicts in how professional interpreters manage their roles in provider-patient interactions. *Social Science & Medicine*, 62(3), 721-730. <https://www.sciencedirect.com/science/article/pii/S027795360500314X>.

Hsieh, E. (2010). Provider-professional interpreter collaboration in bilingual health care: competitions of control over professional interpreter-mediated interactions. *Patient education and counseling*, 78(2), 154-159. <https://www.sciencedirect.com/science/article/pii/S0738399109000925>.

Hsieh, E., & Nicodemus, B. (2015). Conceptualizing emotion in healthcare interpreting: A normative approach to interpreters' emotion work. *Patient education and counseling*, 98(12), 1474-1481. <https://www.sciencedirect.com/science/article/pii/S073839911530001X>

Hsu, M. H. (2024). Mastering medical terminology with ChatGPT and TermBot. *Health Education Journal*, 83(4), 352-358. <https://journals.sagepub.com/doi/abs/10.1177/00178969231197371>.

Huang, W. C., Peloquin, B., Kao, J., Wang, C., Gong, H., Salesky, E., ... & Chen, P. J. (2023, June). A holistic cascade system, benchmark, and human evaluation protocol for expressive speech-to-speech translation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/10096183/>

Hudelson, P., & Chappuis, F. (2024). Using voice-to-voice machine translation to overcome language barriers in clinical communication: an exploratory study. *Journal of General Internal Medicine*, 39(7), 1095-1102. <https://link.springer.com/article/10.1007/s11606-024-08641-w>.

Hussein, A., Yan, B., Anastasopoulos, A., Watanabe, S., & Khudanpur, S. (2024, April). Enhancing end-to-end conversational speech translation through target language context utilization. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 11971-11975). IEEE. <https://ieeexplore.ieee.org/abstract/document/10446102/>

Hutchins, J. (2003). Machine translation. *Machine translation*, 1059-1066. http://www.murieltranslations.com/articles/machine_translation/what_do_we_want-nirenberg.pdf

Iafrate, F. (2014). A journey from big data to smart data. In *Digital Enterprise Design & Management: Proceedings of the Second International Conference on Digital Enterprise Design and Management DED&M 2014* (pp. 25-33). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-319-04313-5_3

IBM. What is machine learning pipeline? (2024). <https://www.ibm.com/think/topics/machine-learning-pipeline>.

Interactio. (2021). Top 8 interpretation failures that prove human interpreting is here for a reason. <https://www.interactio.io/post/top-interpretation-failures-that-prove-human-interpreting-is-here-for-a-reason#:~:text=Facebook%20post%20incites%20violence>

Jacobsen, B. (2004). Pragmatic meaning in court interpreting: An empirical study of additions in consecutively-interpreted question-answer dialogues. *The International Journal of Speech, Language and the Law*, 11(1), 00159-172. <https://utppublishing.com/doi/pdf/10.1558/ijsl.v11i1.165>.

Jakobsson, L. (2016). (im) Politeness and Honorification, in Japanese Translation. <https://lup.lub.lu.se/student-papers/search/publication/8891161>

Jay, R. (2024). Your First Agent App. In *Generative AI Apps with LangChain and Python: A Project-Based Approach to Building Real-World LLM Apps* (pp. 315-344). Berkeley, CA: Apress. https://link.springer.com/chapter/10.1007/979-8-8688-0882-1_8.

Jia, Y., Ramanovich, M. T., Remez, T., & Pomerantz, R. (2022, June). Translatotron 2: High-quality direct speech-to-speech translation with voice preservation. In *International Conference on Machine Learning* (pp. 10120-10134). PMLR. <https://proceedings.mlr.press/v162/jia22b.html>

Jia, Y., Weiss, R. J., Biadsky, F., Macherey, W., Johnson, M., Chen, Z., & Wu, Y. (2019). Direct speech-to-speech translation with a sequence-to-sequence model. *arXiv preprint arXiv:1904.06037*. <https://arxiv.org/abs/1904.06037>.

Jiménez-Crespo, M. A. (2015). Collaborative and volunteer translation and interpreting. In *Researching translation and interpreting* (pp. 58-70). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315707280-7/collaborative-volunteer-translation-interpreting-miguel-jim%C3%A9nez-crespo>.

Jones, R. (2014). *Conference interpreting explained*. Routledge. <https://www.taylorfrancis.com/books/mono/10.4324/9781315760322/conference-interpreting-explained-roderick-jones-anthony-pym>.

Jutorán, M. O. (2023). Dealing with legal terminology in court interpreting. In *Handbook of Terminology* (pp. 570-593). John Benjamins. <https://www.jbe-platform.com/content/books/9789027249388-hot.3.deal>

Kabir, M., Ahmed, T., Rahman, M. M., Giannouris, P., & Ananiadou, S. (2025). Semantic Label Drift in Cross-Cultural Translation. *arXiv preprint arXiv:2510.25967*. <https://arxiv.org/abs/2510.25967>

Kadrić, M., Rennert, S., & Schäffner, C. (2021). Interpreting in politics and diplomacy. In *Diplomatic and Political Interpreting Explained* (pp. 18-41). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780367809898-2/interpreting-politics-diplomacy-mira-kadri%C4%87-sylvi-rennert-christina-sch%C3%A4ffner>.

Kala, J. R., Adetiba, E., Abayom, A., Dare, O. E., & Ifijeh, A. H. (2025). Speech to Speech Translation with Translatotron: A State of the Art Review. *arXiv preprint arXiv:2502.05980*. <https://arxiv.org/abs/2502.05980>.

Kalai, A. T., & Vempala, S. S. (2024, June). Calibrated language models must hallucinate. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing* (pp. 160-171). <https://dl.acm.org/doi/abs/10.1145/3618260.3649777>

- Kalina, S. (2015). Ethical challenges in different interpreting settings. *MonTI: Monografias de Traducción e Interpretación*, 63-86. <https://www.raco.cat/index.php/MonTI/article/view/312880>.
- Kano, T., Sakti, S., & Nakamura, S. (2018). Structured-based curriculum learning for end-to-end english-japanese speech translation. *arXiv preprint arXiv:1802.06003*. <https://arxiv.org/abs/1802.06003>.
- Karaman, B. İ. (2024). Advancing Crisis Translation And Interpreting: The Design Of The Multilingual Smartphone Application CTermPort For Crisis Settings. https://www.preprints.org/manuscript/202402.0073/download/final_file
- Kardas, M., Czapla, P., Stenetorp, P., Ruder, S., Riedel, S., Taylor, R., & Stojnic, R. (2020). Axccl: Automatic extraction of results from machine learning papers. *arXiv preprint arXiv:2004.14356*. <https://arxiv.org/abs/2004.14356>.
- Karttunen, F. E. (1994). *Between worlds: Professional interpreters, guides, and survivors*. Rutgers University Press.
- Karunya, S., Jalakandeshwaran, M., & Uma, R. (2023, December). AI-Powered Real-Time Speech-to-Speech Translation for Virtual Meetings Using Machine Learning Models. In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/10448600/>
- Keidar, D., Opedal, A., Jin, Z., & Sachan, M. (2022). Slangvolution: A causal analysis of semantic change and frequency dynamics in slang. *arXiv preprint arXiv:2203.04651*. <https://arxiv.org/abs/2203.04651>
- Keiser, W. (1999). L'Histoire de l'Association Internationale des Interprètes de Conférence (AIIC). *Interpreting*, 4(1), 81-95. <https://www.erudit.org/en/journals/meta/2004-v49-n3-meta816/009380ar/abstract/>.
- Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron*, 100(2), 424-435. [https://www.cell.com/neuron/fulltext/S0896-6273\(18\)30857-2](https://www.cell.com/neuron/fulltext/S0896-6273(18)30857-2)
- Kenny, D. (2018). Machine translation. In *The Routledge handbook of translation and philosophy* (pp. 428-445). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315678481-27/machine-translation-dorothy-kenny>
- Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 100145. <https://www.sciencedirect.com/science/article/pii/S2666990024000120>.
- Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. *Future Generation Computer Systems*, 97, 219-235. <https://www.sciencedirect.com/science/article/pii/S0167739X18319903>
- Khanam, F., Munmun, F. A., Ritu, N. A., Saha, A. K., & Firoz, M. (2022). Text to speech synthesis: A systematic review, deep learning based architecture and future research direction. *Journal of Advances in Information Technology*, 13(5). <https://www.academia.edu/download/95499988/20220831054604906.pdf>
- Kheddar, H., Himeur, Y., Al-Maadeed, S., Amira, A., & Bensaali, F. (2023). Deep transfer learning for automatic speech recognition: Towards better generalization. *Knowledge-Based Systems*, 277, 110851. <https://www.sciencedirect.com/science/article/pii/S0950705123006019>

- Kiktova, E., Zimmermann, J., Pařová, M., & Sock, R. (2020). Detecting Anticipation Foci for Simultaneous Interpreting. *SKASE Journal of Theoretical Linguistics*, 17(4). http://www.skase.sk/Volumes/JTL46/pdf_doc/04.pdf
- Kim, J. H., Kim, J., Park, J., Kim, C., Jhang, J., & King, B. (2025). When ChatGPT gives incorrect answers: the impact of inaccurate information by generative AI on tourism decision-making. *Journal of Travel Research*, 64(1), 51-73. <https://journals.sagepub.com/doi/abs/10.1177/00472875231212996>.
- Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. In *Workshop on Post-Editing Technology and Practice*. <https://aclanthology.org/2012.amta-wptp.2/>
- Korpala, P., & Jasielska, A. (2019). Investigating professional interpreters' empathy: Are emotions in simultaneous interpreting contagious? *Target*, 31(1), 2-24. <https://www.jbe-platform.com/content/journals/10.1075/target.17123.kor>.
- Kroeger, P. R. (2023). *Analyzing meaning: An introduction to semantics and pragmatics*. Language Science Press. <https://library.oapen.org/handle/20.500.12657/61714>.
- Kudo. (2025). AI Speech Translation in 2025 & Beyond: Technology, Data, Trends & Predictions. <https://kudo.ai/blog/ai-speech-translation-in-2025-beyond-technology-data-trends-predictions/#:~:text=,and%20Cultural%20Context>
- Kumar, G., Post, M., Povey, D., & Khudanpur, S. (2014, May). Some insights from translating conversational telephone speech. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3231-3235). IEEE. Available in: <https://ieeexplore.ieee.org/abstract/document/6854197/>.
- Kumar, S., Thakur, J., Ekka, D., & Sahu, I. (2022). Web scraping using Python. *International Journal of Advances in Engineering and Management*, 4(9), 235-237. https://ijaem.net/issue_dcp/Web%20Scraping%20Using%20Python.pdf.
- Kurz, I. (1985). The rock tombs of the Princes of Elephantine. *Babel*, 31(4), 213-218. <https://open.unive.it/hitrade/books/KurzRock.pdf>.
- Kushalnagar, P., Paludneviciene, R., & Kushalnagar, R. (2019). Video remote interpreting technology in health care: Cross-sectional study of deaf patients' experiences. *JMIR Rehabilitation and Assistive Technologies*, 6(1), e13233. <https://rehab.jmir.org/2019/1/e13233>.
- Lang, C. A. (2023). *Identity Construction and Role in Interpreted Interactions*. Gallaudet University. <https://search.proquest.com/openview/63117ec6abd6524d17d6cd772e284424/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- Lee, A., Chen, P. J., Wang, C., Gu, J., Popuri, S., Ma, X., ... & Hsu, W. N. (2021). Direct speech-to-speech translation with discrete units. *arXiv preprint arXiv:2107.05604*. <https://arxiv.org/abs/2107.05604>
- Lee, J. (2020, November). Professional identity development among student professional interpreters: A case study. In *Forum* (Vol. 18, No. 2, pp. 179-196). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://www.jbe-platform.com/content/journals/10.1075/forum.20003.lee>.
- Lee, S., & Kim, H. J. (2008, September). News keyword extraction for topic tracking. In *2008 fourth international conference on networked computing and advanced information management* (Vol. 2, pp. 554-559). IEEE. <https://ieeexplore.ieee.org/abstract/document/4624203/>.

- Lei, X. (2024). Real-time translation of English speech through speech feature extraction. *Artificial Life and Robotics*, 29(3), 410-415. <https://link.springer.com/article/10.1007/s10015-024-00951-w>.
- Lewis, W. D. (2015). Skype translator: Breaking down language and hearing barriers. *Translating and the Computer (TC37)*, 10, 125-149. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/TC37-Paper-FINAL.pdf>
- Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(4), 745-777. <https://ieeexplore.ieee.org/abstract/document/6732927/>
- Li, X., Jia, Y., & Chiu, C. C. (2023, June). Textless direct speech-to-speech translation with discrete speech representation. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/10096797/>.
- Li, X., Metsis, V., Wang, H., & Ngu, A. H. H. (2022, June). Tts-gan: A transformer-based time-series generative adversarial network. In *International conference on artificial intelligence in medicine* (pp. 133-143). Cham: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-09342-5_13.
- Lin, Y. C., Lin, T. Q., Yang, C. K., Lu, K. H., Chen, W. C., Kuan, C. Y., & Lee, H. Y. (2024, December). Listen and Speak Fairly: a Study on Semantic Gender Bias in Speech Integrated Large Language Models. In *2024 IEEE Spoken Language Technology Workshop (SLT)* (pp. 439-446). IEEE. <https://ieeexplore.ieee.org/abstract/document/10832317/>
- LingoAid. (2024, March 22). *The 9 essential skills for simultaneous interpretation mastery*. LinkedIn. <https://www.linkedin.com/pulse/9-essential-skills-simultaneous-interpretation-mastery-lingoaid-ujamc>. Retrieved 28 March 2025.
- Littig, B., & Pöchhacker, F. (2014). Socio-translational collaboration in qualitative inquiry: The case of expert interviews. *Qualitative inquiry*, 20(9), 1085-1095. <https://journals.sagepub.com/doi/abs/10.1177/1077800414543696>.
- Liu, D., Wang, C., Gong, H., Ma, X., Tang, Y., & Pino, J. (2021). From Start to Finish: Latency Reduction Strategies for Incremental Speech Synthesis in Simultaneous Speech-to-Speech Translation. *arXiv preprint arXiv:2110.08214*. <https://arxiv.org/abs/2110.08214>
- Llewellyn-Jones, P., & Lee, R. G. (2014). *Redefining the role of the community professional interpreter: The concept of role-space*. Sli Press. https://clouk.uclan.ac.uk/10818/?template=default_internal.
- Lloyd, K. (2018). Bias amplification in artificial intelligence systems. *arXiv preprint arXiv:1809.07842*. <https://arxiv.org/abs/1809.07842>
- Loiseau, N., & Delgado Luchner, C. (2021). A, B and C decoded: understanding professional interpreters' language combinations in terms of language proficiency. *The Professional interpreter and Translator Trainer*, 15(4), 468-489. <https://www.tandfonline.com/doi/abs/10.1080/1750399X.2021.1911193>.
- Longley, P. (1984). The International Association of Conference Professional interpreters (AIIC). <https://www.degruyter.com/document/doi/10.1515/mult.1984.3.1.31/html>.
- Luo, C. (2024). Has llm reached the scaling ceiling yet? unified insights into llm regularities and constraints. *arXiv preprint arXiv:2412.16443*. <https://arxiv.org/abs/2412.16443>.

Ma, M., Zheng, B., Liu, K., Zheng, R., Liu, H., Peng, K., ... & Huang, L. (2019). Incremental text-to-speech synthesis with prefix-to-prefix framework. *arXiv preprint arXiv:1911.02750*. <https://arxiv.org/abs/1911.02750>,

Macaulay, T. (2023, April). Swiss startup unveils ‘world-first’ AI translation service: A breakthrough for automated interpretations? The Next Web. <https://thenextweb.com/news/interprefy-startup-unveils-worlds-first-advanced-automated-speech-translation-service-for-online-and-live-eventsinterprefy#:~:text=Interprefy%20supplies%20a%20way%20to,at%20the%20International%20S pace%20Station>

Macnamara, B. N., Moore, A. B., Kegl, J. A., & Conway, A. R. (2014). Domain-general cognitive abilities and simultaneous interpreting skill. In *Aptitude for interpreting* (pp. 107-128). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/bct.68.06mac/html>.

Mairs, R. (2011). Translator, traditor: The professional interpreter as traitor in classical tradition. *Greece & Rome*, 58(1), 64-81. <https://www.cambridge.org/core/journals/greece-and-rome/article/translator-traditor-the-professional-interpreter-as-traitor-in-classical-tradition/2C04DD749795D172216FAF42A1CE3F86>.

Marais, L., Louw, J. A., Badenhorst, J., Calteaux, K., Wilken, I., Van Niekerk, N., & Stein, G. (2020, July). AwezaMed: A multilingual, multimodal speech-to-speech translation application for maternal health care. In 2020 IEEE 23rd International Conference on Information Fusion (FUSION) (pp. 1-8). IEEE. <https://ieeexplore.ieee.org/abstract/document/9190240/>

Marangoz, E. Ö. (2023). Balancing Ethical Responsibility and Fairness: Considerations in Conference Interpreting. *Journal of Academic Social Science Studies*, 16(96). <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=21472971&AN=173678046&h=9TFqTo9xM%2FUF%2FfwGJ%2BRhCkytZsyZ1xSKRI6rTcJVYRe7BW6nHZ1Rhbaa6pauP2igjyliWMKXMzRZEHjLCiLybA%3D%3D&crl=c>.

Marecka, M., Szewczyk, J., Otwinowska, A., Durlík, J., Foryś-Nogala, M., Kutylowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition*, 206, 104477. <https://www.sciencedirect.com/science/article/pii/S0010027720302961>.

Martínez-Gómez, A. (2015). Invisible, visible or everywhere in between? Perceptions and actual behaviors of non-professional professional interpreters and interpreting users. <https://www.openstarts.units.it/bitstreams/66402c5d-614d-45c5-8622-f0dde59209e2/download>.

Mason, I., & Ren, W. (2014). Power in face-to-face interpreting events. In *The sociological turn in translation and interpreting studies* (pp. 115-133). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/bct.66.08mas/html>.

Mathur, N., Baldwin, T., & Cohn, T. (2020). Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. *arXiv preprint arXiv:2006.06264*. <https://arxiv.org/abs/2006.06264>

McDonald, J. L., & Carpenter, P. A. (1981). Simultaneous translation: Idiom interpretation and parsing heuristics. *Journal of Verbal Learning and Verbal Behavior*, 20(2), 231-247. <https://www.sciencedirect.com/science/article/pii/S0022537181903972>.

McGiff, J., & Nikolov, N. S. (2025). Overcoming Data Scarcity in Generative Language Modelling for Low-Resource Languages: A Systematic Review. *arXiv preprint arXiv:2505.04531*. <https://arxiv.org/abs/2505.04531>

Mechkaroska, D., Domazet, E., Feta, A., & Shikoska, U. R. (2024, March). Architectural scalability of conversational chatbot: the case of ChatGPT. In *Future of Information and Communication Conference* (pp. 54-71). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-53960-2_5.

Medhi, I., Patnaik, S., Brunskill, E., Gautama, S. N., Thies, W., & Toyama, K. (2011). Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(1), 1-28. <https://dl.acm.org/doi/abs/10.1145/1959022.1959024>

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6), 1-35. <https://dl.acm.org/doi/abs/10.1145/3457607>

Metzger, M. A. (1996). *The paradox of neutrality: A comparison of professional interpreters' goals with the reality of interactive discourse*. Georgetown University. <https://search.proquest.com/openview/398e49dbc6abb6b9e7f37ba9cae7ad3a/1?pq-origsite=gscholar&cbl=18750&diss=y>.

Metzger, M. (1999). *Sign language interpreting: Deconstructing the myth of neutrality*. Gallaudet University Press. <https://books.google.com.br/books?hl=pt-BR&lr=&id=cF2FLg0fEpEC&oi=fnd&pg=PA1&dq=Metzger+1999&ots=EEeIeLlKAw&sig=AXdntAxHBG9TQBCNJKxfM4zqWpE>.

Mikkelsen, H. (1996, October). The professionalization of community interpreting. In *Global vision: Proceedings of the 37th annual conference of the American Translators Association* (pp. 77-89). https://aiic.org/document/92/AIICBlog_Oct2004_MIKKELSON_Community_interpreting_EN.pdf.

Mikkelsen, H. (1999, May). Court interpreting at a crossroads. In *Proceedings of Annual Conference of the National Association of Judiciary Professional interpreters and Translators (NAJIT)*.—Washington. <https://acebo.myshopify.com/pages/court-interpreting-at-a-crossroads>.

Mikkelsen, H. (2000). Professional interpreter ethics: A review of the traditional and electronic literature. *Interpreting*, 5(1), 49-56. <https://www.jbe-platform.com/content/journals/10.1075/intp.5.1.05mik>.

Mikkelsen, H. (2008). Evolving views of the court professional interpreter's role: Between Scylla and Charybdis. In *Crossing borders in community interpreting: Definitions and dilemmas* (pp. 81-97). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.76.05mik/html>.

Mikkelsen, H., & Jourdenais, R. (Eds.). (2015). *The Routledge handbook of interpreting* (p. 456). London and New York: Routledge. <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315745381&type=googlepdf>.

Mills, A., & Angell, N. (2025). Are We Tripping? The Mirage of AI Hallucinations. *The Mirage of AI Hallucinations* (February 06, 2025). <https://dx.doi.org/10.2139/ssrn.5127162>

Min, A., Hu, C., Ren, Y., & Zhao, H. (2025). A Unit-based System and Dataset for Expressive Direct Speech-to-Speech Translation. *arXiv preprint arXiv:2502.00374*. <https://arxiv.org/abs/2502.00374>.

Min, A., Hu, C., Ren, Y., & Zhao, H. (2025). When End-to-End is Overkill: Rethinking Cascaded Speech-to-Text Translation. *arXiv preprint arXiv:2502.00377*. <https://arxiv.org/abs/2502.00377>

- Mondal, H., & Mondal, S. (2024). Artificial intelligence-generated content needs a human oversight. *Indian Journal of Dermatology*, 69(3), 284. https://journals.lww.com/ijd/fulltext/2024/05000/artificial_intelligence_generated_content_needs_a.30.aspx?context=latestarticles.
- Moratto, R., & Yang, Z. (2024). Probing the cognitive load of consecutive professional interpreters: A corpus-based study. *Translation and Interpreting Studies*, 19(2), 234-256. <https://www.jbe-platform.com/content/journals/10.1075/tis.22047.mor>.
- Morris, R. (1997). Technology and the Worlds of Interpreting. *Future and communication: The role of scientific and technical communication and translation in technology development and transfer*, 177-184. <https://open.unive.it/hitrade/books/Technology.pdf>.
- Moser-Mercer, B. (2005). Remote interpreting: Issues of multi-sensory integration in a multilingual task. *Meta*, 50(2), 727-738. <https://www.erudit.org/en/journals/meta/2005-v50-n2-meta881/011014ar.pdf>.
- Mozziconacci, S. J., & Hermes, D. J. (1999). Role of intonation patterns in conveying emotion in speech. In *Proceedings of the 14th International congress of phonetic sciences, ICPHS-99, San Francisco, CA, USA, August 1-7, 1999* (pp. 2001-2004). <https://research.tue.nl/files/1354072/696624203788951.pdf>
- Mujadia, V., & Sharma, D. M. (2023, May). Towards speech to speech machine translation focusing on indian languages. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 161-168). <https://aclanthology.org/2023.eacl-demo.19/>.
- Munane, F. A. (2014). *Cultural taboos as a factor in interpretation in the medical field. The case study of the Meridian hospital* (Doctoral dissertation, University of Nairobi). <https://erepository.uonbi.ac.ke/handle/11295/75992>.
- Nachman, J. (2020). *Linguistic heroes or necessary evil?* Community interpreting and the phenomenon of language brokering in bilingual schools in Poland and Polish schools abroad. <https://ruj.uj.edu.pl/entities/publication/5ff0cafb-7b00-4aca-a837-35466fec2966>.
- Nakamura, S., Markov, K., Nakaiwa, H., Kikui, G. I., Kawai, H., Jitsuhiro, T., ... & Yamamoto, S. (2006). The ATR multilingual speech-to-speech translation system. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(2), 365-376. <https://ieeexplore.ieee.org/abstract/document/1597243/>.
- Nakane, I. (2009). The myth of an'invisible mediator': An Australian case study of English-Japanese police interpreting. *Portal: Journal of multidisciplinary international studies*, 6(1), 1-16. <https://search.informit.org/doi/abs/10.3316/informit.074729605752103>.
- National Council on Interpreting in Health Care. (2021). Professional interpreter advocacy in healthcare encounters: A closer look. *National Council on Interpreting in Health Care*. https://www.ncihc.org/assets/z2021Images/Professional_interpreter%20Advocacy%20in%20Healthcare%20Encounters%20A%20Closer%20Look%20F051121.pdf
- Nguyen, H., Bougares, F., Tomashenko, N., Esteve, Y., & Besacier, L. (2020, October). Investigating self-supervised pre-training for end-to-end speech translation. In *Interspeech 2020*. <https://hal.science/hal-02962186/>

Nick, S. (2001). Use of language in diplomacy. *Language and Diplomacy*. Malta: DiploProjects, 39-48. https://books.google.com.br/books?hl=pt-BR&lr=&id=yKcHHU2DaPoC&oi=fnd&pg=PA39&dq=Use+of+Language+in+Diplomacy&ots=ddwEafV_u1&sig=Y_VLdfJg_rrSzP60hH_CriuZF6Y.

Nishat, A., & Mustafa, A. (2024). Domain-Specific Fine-Tuning of BERT and ChatGPT for Enhanced Medical Text Analysis. *Journal of Computational Innovation*, 4(1). <https://researchworkx.com/index.php/jci/article/view/44>.

Niska, H. (2008). Community professional interpreter training: Past, present, future. In *Interpreting in the 21st Century: Challenges and opportunities* (pp. 133-144). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.43.14nis/html>.

Och, F. J., & Ney, H. (2000). Statistical machine translation. In *5th EAMT Workshop: Harvesting Existing Resources*. <https://aclanthology.org/2000.eamt-1.5.pdf>

Okoniewska, A. M. (2022). Professional interpreters' roles in a changing environment. *The Translator*, 28(2), 139-147. <https://doi.org/10.1080/13556509.2022.2124712>

Ono, N., & Yang, J. (2024, October). Communication skills of medical professional interpreters: a qualitative explanatory study of healthcare professionals' and medical professional interpreters' perceptions. In *Healthcare* (Vol. 12, No. 20, p. 2073). MDPI. <https://www.mdpi.com/2227-9032/12/20/2073>

Onyshchak, H., Liutko, N., Yarova, A., Povoroznyuk, R., Kolomiets, I., & Gontsa, I. (2023). Pragmatic competence in political discourse interpreting. *Revista Romaneasca Pentru Educatie Multidimensionala*, 15(3), 376-399. <https://lumenpublishing.com/journals/index.php/rrem/article/view/6114>

Ouyang, S., Hrinchuk, O., Chen, Z., Lavrukhin, V., Balam, J., Li, L., & Ginsburg, B. (2025, April). Anticipating Future with Large Language Model for Simultaneous Machine Translation. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 5547-5557). <https://aclanthology.org/2025.naacl-long.286/>

Paez, N. (2018, November). Versioning strategy for DevOps implementations. In *2018 Congreso Argentino de Ciencias de La Informática y Desarrollos de Investigación (CACIDI)* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/8584362/>

Pal, S. (2023). A paradigm shift in research: Exploring the intersection of artificial intelligence and research methodology. *International journal of innovative research in engineering & multidisciplinary physical sciences*, 11(3), 1-7. <https://pdfs.semanticscholar.org/3f46/659fe79a4a86f0580dd5529a49d05cca6c69.pdf>.

Pamadi, V. N., & Singh, P. (2025). Edge AI vs Cloud AI: A Comparative Study of Performance Latency and Scalability. *International Journal of Research in Modern Engineering & Emerging Technology (IJRMEET)*, 13(3), 13-35. <https://elibrary.ru/item.asp?id=81773551>

Pamisetty, G., & Sri Rama Murty, K. (2023). Prosody-TTS: An end-to-end speech synthesis system with prosody control. *Circuits, Systems, and Signal Processing*, 42(1), 361-384. <https://link.springer.com/article/10.1007/s00034-022-02126-z>

Parenzo, S., & Schuster, M. (2020). The mental health professional interpreter: The “third space” between transference and counter-transference. In *Handbook of Research on Medical Interpreting* (pp.

276-289). IGI Global. <https://www.igi-global.com/chapter/the-mental-health-professional-interpreter/246126>.

Paskin, N. (2010). Digital object identifier (DOI®) system. *Encyclopedia of library and information sciences*, 3, 1586-1592. <https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.1081/E-ELIS4-133&type=chapterpdf>

Pepiler, C. W. (1915). *The Professional interpreters of Foreign Languages among the Ancients: A Study Based on Greek and Latin Sources*. <https://www.jstor.org/stable/4387125>.

Phelan, M. (2019). Codes of ethics. In *Ethics in public service interpreting* (pp. 85-146). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315715056-3/codes-ethics-mary-phelan>.

Piller, I. (2021, June). The interpreting profession in Ancient Egypt. *Language on the move*. <https://www.languageonthemove.com/the-interpreting-profession-in-ancient-egypt>. Retrieve 23 March 2025.

Placani, A. (2024). Anthropomorphism in AI: hype and fallacy. *AI and Ethics*, 4(3), 691-698. <https://link.springer.com/article/10.1007/s43681-024-00419-4>

Platnick, D., Abdelnour, B., Earl, E., Kumar, R., Rezaei, Z., Tsangaris, T., & Lagum, F. (2024). Preset-Voice Matching for Privacy Regulated Speech-to-Speech Translation Systems. *arXiv preprint arXiv:2407.13153*. <https://arxiv.org/abs/2407.13153>

Pöchhacker, F. (2006). Research and methodology in healthcare interpreting. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 5. <https://lans.uantwerpen.be/index.php/LANS-TTS/article/view/157>.

Pöchhacker, F. (2008). Researching interpreting quality: Models and methods. In *Interpreting in the 21st century: Challenges and opportunities* (pp. 95-106). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.43.10poc/html>.

Pöchhacker, F. (2013). Teaching interpreting/Training professional interpreters. In *Handbook of Translation Studies: Volume 4* (pp. 174-180). John Benjamins Publishing Company. <https://doi.org/10.1075/hts.4.tea2>.

Pöchhacker, F. (2022). *Introducing interpreting studies*. Routledge. <https://www.taylorfrancis.com/books/mono/10.4324/9781003186472/introducing-interpreting-studies-franz-p%C3%B6chhacker>.

Pöchhacker, F. (2024). Is machine interpreting interpreting?. *Translation spaces*. <https://doi.org/10.1075/ts.23028.poc>

Polák, P., & Bojar, O. (2024, December). Long-form end-to-end speech translation via latent alignment segmentation. In *2024 IEEE Spoken Language Technology Workshop (SLT)* (pp. 1076-1082). IEEE. <https://ieeexplore.ieee.org/abstract/document/10832264/>

Prabhavalkar, R., Hori, T., Sainath, T. N., Schlüter, R., & Watanabe, S. (2023). End-to-end speech recognition: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 325-351. <https://ieeexplore.ieee.org/abstract/document/10301513/>

Prajwal, K. R., Hegde, S., & Zisserman, A. (2025, April). Scaling Multilingual Visual Speech Recognition. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/10890395/>

Prasad, R., Kumar, R., Ananthakrishnan, S., Chen, W., Hewavitharana, S., Roy, M., ... & Natarajan, P. (2012). Active error detection and resolution for speech-to-speech translation. In *Proceedings of the 9th International Workshop on Spoken Language Translation: Papers* (pp. 150-157). <https://aclanthology.org/2012.iwslt-papers.1/>.

Prenger, R., Valle, R., & Catanzaro, B. (2019, May). Waveglow: A flow-based generative network for speech synthesis. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3617-3621). IEEE. <https://ieeexplore.ieee.org/abstract/document/8683143/>.

Propio. (2024) How Propio Handles Bias in Machine Translations. <https://propio.com/2024/02/21/how-propio-handles-bias-in-machine-translations/#:~:text=Bias%20in%20neural%20machine%20translation,gender%20representation%2C%20and%20racial%20sensitivity>

Raganato, A., & Tiedemann, J. (2018, November). An analysis of encoder representations in transformer-based machine translation. In *Proceedings of the 2018 EMNLP workshop BlackboxNLP: analyzing and interpreting neural networks for NLP* (pp. 287-297). <https://aclanthology.org/W18-5431/>.

Ramos, F. P. (2015). Institutional Translation and Interpreting. *Focus*, 23, 27. <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780429264894&type=googlepdf>.

Ranathunga, S., Lee, E. S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R. (2023). Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys*, 55(11), 1-37. <https://dl.acm.org/doi/abs/10.1145/3567592>

Rashidi, S., & Sameti, H. (2025). Improving Direct Persian-English Speech-to-Speech Translation with Discrete Units and Synthetic Parallel Data. *arXiv preprint arXiv:2511.12690*. <https://arxiv.org/abs/2511.12690>

Reiswig, J. (2010). Mendeley. *Journal of the Medical Library Association: JMLA*, 98(2), 193. <https://pmc.ncbi.nlm.nih.gov/articles/PMC2859264/>

Robertson, S., & Díaz, M. (2022, June). Understanding and being understood: User strategies for identifying and recovering from mistranslations in machine translation-mediated chat. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2223-2238). <https://dl.acm.org/doi/abs/10.1145/3531146.3534638>

Roobahani, Z. (2025). A Review of Methods for Reducing Hallucinations in Generative Artificial Intelligence to Enhance Knowledge Economy. *Knowledge Economy Studies*.

Rosenberg, E., Seller, R., & Leanza, Y. (2008). Through professional interpreters' eyes: comparing roles of professional and family professional interpreters. *Patient education and counseling*, 70(1), 87-93. <https://www.sciencedirect.com/science/article/pii/S0738399107003771>.

Rudvin, M. (2008). Negotiating linguistic and cultural identities in professional interpreter-mediated communication for public health services. In *Sociocultural aspects of translating and interpreting* (pp. 173-190). John Benjamins Publishing Company. <https://www.degruyter.com/document/doi/10.1075/btl.67.22rud/html>.

Rudvin, M. (2015). Interpreting and professional identity. In *The Routledge handbook of interpreting* (pp. 432-446). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315745381-32/interpreting-professional-identity-mette-rudvin>.

Ruffolo, J. (2023, June 6). *Subject matter expertise in translation services*. JR Language Translation Services. <https://blog-english.jrlanguage.com/subject-matter-expertise-in-translation-services/>. Retrieved 28 March 2025.

Russell, D., & Takeda, K. (2015). Consecutive interpreting. In *The Routledge handbook of interpreting* (pp. 96-109). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315745381-9/consecutive-interpreting-debra-russell-kayoko-takeda>.

Saabith, S., Vinothraj, T., & Fareez, M. (2021). A review on Python libraries and Ides for Data Science. *Int. J. Res. Eng. Sci*, 9(11), 36-53. https://www.researchgate.net/profile/Vinothraj-Thangarajah/publication/357898994_A_Review_on_Python_Libraries_and_IDEs_for_Data_Science/inks/620249344d89183b338b49c2/A-Review-on-Python-Libraries-and-IDEs-for-Data-Science.pdf.

Saadany, H., & Orasan, C. (2021). BLEU, METEOR, BERTScore: Evaluation of metrics performance in assessing critical translation errors in sentiment-oriented text. *arXiv preprint arXiv:2109.14250*. <https://arxiv.org/abs/2109.14250>

Safar, D., Safar, M., Jaafar, S., Al-Yachli, B. A., Shukri, A. K., & Rasheed, M. H. (2025). Hallucinations in GPT-2 Trained Model. *Ingénierie des Systèmes d'Information*, 30(1). <https://ieta.org/journals/isi/paper/10.18280/isi.300104>

Saina, F. (2021, July). Remote interpreting: Platform testing in a university setting. In *Proceedings of the Translation and Interpreting Technology Online Conference* (pp. 57-67). Available in: <https://aclanthology.org/2021.triton-1.7/>.

Salazar, J., Kirchhoff, K., & Huang, Z. (2019, May). Self-attention networks for connectionist temporal classification in speech recognition. In *Icassp 2019-2019 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 7115-7119). IEEE. <https://ieeexplore.ieee.org/abstract/document/8682539/>

Sarim, M., Shakeel, S., Javed, L., & Nadeem, M. (2025). Direct Speech to Speech Translation: A Review. *arXiv preprint arXiv:2503.04799*. <https://arxiv.org/abs/2503.04799>

Saunders, D. (2022). Domain adaptation and multi-domain adaptation for neural machine translation: A survey. *Journal of Artificial Intelligence Research*, 75, 351-424.

Schlögl, S., Doherty, G., & Luz, S. (2015). Wizard of Oz experimentation for language technology applications: Challenges and tools. *Interacting with Computers*, 27(6), 592-615. <https://arxiv.org/html/2402.14563v1#:~:text=Wizard%20of%20OZ%20,in%20such%20cases%20achieving%20acceptable>

Schmidt, C. W., Reddy, V., Zhang, H., Alameddine, A., Uzan, O., Pinter, Y., & Tanner, C. (2024). Tokenization is more than compression. *arXiv preprint arXiv:2402.18376*. <https://arxiv.org/abs/2402.18376>.

Seeber, K. G. (2013). Cognitive load in simultaneous interpreting: Measures and methods. *Target. International Journal of Translation Studies*, 25(1), 18-32.

Seligman, M., & Dillinger, M. (2006, June). Usability issues in an interactive speech-to-speech translation system for healthcare. In *Proceedings of the First International Workshop on Medical Speech Translation* (pp. 1-4). <https://aclanthology.org/W06-3701.pdf>

Setton, R., & Dawrant, A. (2016). Conference interpreting: A complete course. <https://www.torrossa.com/gs/resourceProxy?an=5001889&publisher=FZ4850>.

Sharrab, Y. O., Attar, H., Eljinini, M. A. H., Al-Omary, Y., & Al-Momani, W. A. (2025). Advancements in Speech Recognition: A Systematic Review of Deep Learning Transformer Models, Trends, Innovations, and Future Directions. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/10924161/>

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4779-4783). IEEE. <https://ieeexplore.ieee.org/abstract/document/8461368/>.

Sheneman, N. (2016). *Deaf professional interpreters' ethics*: Reflections on training and decision-making. *Journal of Interpretation*, 25(1), 8. <https://digitalcommons.unf.edu/joi/vol25/iss1/8/>.

Shiwen, Y., & Xiaojing, B. (2014). Rule-based machine translation. In *Routledge encyclopedia of translation technology* (pp. 186-200). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315749129-11/rule-based-machine-translation-yu-shiwen-bai-xiaojing>

Singh, J., & Singh, D. (2024). A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery: Taxonomy, challenges, applications and future prospects. *Advanced Engineering Informatics*, 62, 102799. <https://www.sciencedirect.com/science/article/pii/S1474034624004476>.

Smeets, R. (2004). Language as a vehicle of the intangible cultural heritage. *Museum international*, 56(1-2), 156-165. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1350-0775.2004.00470.x>.

Smith, S. L. An exercise to practise salami technique. The Interpreting Coach. https://www.theinterpretingcoach.com/an-exercise-to-practise-salami-technique/?utm_source=chatgpt.com

Soni, A. A. (2025). *Improving speech recognition accuracy using custom language models with the Vosk Toolkit* (arXiv preprint arXiv:2503.21025). arXiv. <https://arxiv.org/abs/2503.21025>

Spechbach, H., Gerlach, J., Karker, S. M., Tsourakis, N., Combescure, C., & Bouillon, P. (2019). A speech-enabled fixed-phrase translator for emergency settings: Crossover study. *JMIR medical informatics*, 7(2), e13167. <https://pmc.ncbi.nlm.nih.gov/articles/PMC6528434/#:~:text=In%20the%20context%20of%20the,but%20instead%20of%20searching%20for>

Spechbach, H., Gerlach, J., Karker, S. M., Tsourakis, N., Combescure, C., & Bouillon, P. (2019). A speech-enabled fixed-phrase translator for emergency settings: Crossover study. *JMIR medical informatics*, 7(2), e13167. <https://medinform.jmir.org/2019/2/e13167/>

Sperber, M., & Paulik, M. (2020). Speech translation and the end-to-end promise: Taking stock of where we are. *arXiv preprint arXiv:2004.06358*. <https://arxiv.org/abs/2004.06358>.

Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69, 343-418. <https://www.jair.org/index.php/jair/article/view/12007>.

Steinfeld, N. (2016). "I agree to the terms and conditions":(How) do users read privacy policies online? An eye-tracking experiment. *Computers in human behavior*, 55, 992-1000. <https://www.sciencedirect.com/science/article/pii/S0747563215301692>

Story, B. H. (2019). History of speech synthesis. In *The Routledge handbook of phonetics* (pp. 9-33). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429056253-2/history-speech-synthesis-brad-story>

Streijl, R. C., Winkler, S., & Hands, D. S. (2016). Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. *Multimedia Systems*, 22(2), 213-227. <https://link.springer.com/article/10.1007/s00530-014-0446-1>

Sun, Z., Zemel, R., & Xu, Y. (2022). Semantically informed slang interpretation. *arXiv preprint arXiv:2205.00616*. <https://arxiv.org/abs/2205.00616>.

Sunyaev, A., & Sunyaev, A. (2020). Cloud computing. *Internet computing: Principles of distributed systems and emerging internet-based technologies*, 195-236. https://link.springer.com/chapter/10.1007/978-3-030-34957-8_7.

Szandała, T. (2025). ChatGPT vs human expertise in the context of IT recruitment. *Expert Systems with Applications*, 264, 125868. <https://www.sciencedirect.com/science/article/pii/S0957417424027350>.

Tang, Y., Sun, A. Y., Inaguma, H., Chen, X., Dong, N., Ma, X., ... & Pino, J. (2023). Hybrid transducer and attention based encoder-decoder modeling for speech-to-text tasks. *arXiv preprint arXiv:2305.03101*. <https://arxiv.org/abs/2305.03101>

Tennant, K., Long, A., & Toney-Butler, T. J. (2017). Active listening. <https://europepmc.org/article/nbk/nbk442015>.

Testoni, A., & Fernández, R. (2024). Asking the right question at the right time: Human and model uncertainty guidance to ask clarification questions. *arXiv preprint arXiv:2402.06509*. <https://arxiv.org/abs/2402.06509>

The National Council on Interpreting in Health Care. (2003, November). Guide to Professional interpreter Positioning in Health Care Settings. *The National Council on Interpreting in Health Care*. <https://www.ncihc.org/assets/documents/workingpapers/NCIHC%20Working%20Paper%20-%20Guide%20to%20Interpreter%20Positioning%20in%20Health%20Care%20Settings.pdf>.

Tiselius, E. (2021, November). Conference and community interpreting: Commonalities and differences. In *The Routledge handbook of conference interpreting* (pp. 49-63). Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9780429297878-6/conference-community-interpreting-elisabet-tiselius>.

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters' Newsletter*, 10(4), 103-127.

Titus, L. M. (2024). Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy. *Cognitive Systems Research*, 83, 101174. <https://www.sciencedirect.com/science/article/pii/S1389041723001080>.

Tjandra, A., Sakti, S., & Nakamura, S. (2019, December). Speech-to-speech translation between untranscribed unknown languages. In *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 593-600). IEEE. <https://ieeexplore.ieee.org/abstract/document/9003853/>.

TransPerfect. (2024, September 28). *Professional interpreter ethics: Navigating confidentiality and impartiality*. TransPerfect Blog. <https://www.transperfect.com/blog/professional-interpreter-ethics-navigating-confidentiality-and-impartiality>. Retrieved 28 March 2025.

Tryuk, M. (2023). La théorie interprétative de Danica Seleskovitch vingt ans après. *Między Oryginałem a Przekładem*, (61), 51-68. <https://www.ceeol.com/search/article-detail?id=1215525>.

Tsiamas, I., Sperber, M., Finch, A., & Garg, S. (2024). Speech is More Than Words: Do Speech-to-Text Translation Systems Leverage Prosody?. *arXiv preprint arXiv:2410.24019*. <https://arxiv.org/abs/2410.24019>

Turobov, A., Coyle, D., & Harding, V. (2024). Using ChatGPT for thematic analysis. *arXiv preprint arXiv:2405.08828*. <https://arxiv.org/abs/2405.08828>.

Uliniansyah, M. T., Riza, H., Santosa, A., Gunawan, M., & Nurfadhilah, E. (2017, November). Development of text and speech corpus for an Indonesian speech-to-speech translation system. In *2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/8384448/>.

Valdeón, R. A. (2013). Doña Marina/La Malinche: A historiographical approach to the professional interpreter/traitor. *Target. International Journal of Translation Studies*, 25(2), 157-179. <https://www.jbe-platform.com/content/journals/10.1075/target.25.2.02val>.

Valero Garcés, C. (ed.) (2003). *Traducción e interpretación en los servicios públicos. Contextualización, actualidad y futuro*. Granada: Editorial Comares. https://www.comares.com/libro/traduccion-e-interpretacion-en-los-servicios-publicos-contextualizacion-actualidad-y-futuro_74858/.

Valero-Garcés, C. (2007). Challenges in multilingual societies. The myth of the invisible professional interpreter and translator. *Across languages and cultures*, 8(1), 81-101. <https://akjournals.com/view/journals/084/8/1/article-p81.xml>.

Valdés, G. (2014). *Expanding definitions of giftedness: The case of young professional interpreters from immigrant communities*. Routledge.

van Rossum, G., & Drake, F. L. (2009). *Python*. https://courses.minia.edu.eg/Attach/16028python_lecture1.pdf.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>

Veluri, B., Itani, M., Chen, T., Yoshioka, T., & Gollakota, S. (2024, May). Look once to hear: Target speech hearing with noisy examples. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). <https://dl.acm.org/doi/abs/10.1145/3613904.3642057>.

Vogler, N., Stewart, C., & Neubig, G. (2019). Lost in interpretation: Predicting untranslated terminology in simultaneous interpretation. *arXiv preprint arXiv:1904.00930*.

von Neumann, T., Boeddeker, C., Kinoshita, K., Delcroix, M., & Haeb-Umbach, R. (2023, June). On word error rate definitions and their efficient computation for multi-speaker speech recognition systems. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/abstract/document/10094784/>

Wadensjö, C. (1992). *Interpreting as interaction: On dialogue-interpreting in immigration hearings and medical encounters* (Doctoral dissertation, Linköpings universitet). <https://www.diva-portal.org/smash/record.jsf?pid=diva2:255952>.

Wadensjö, C. (1998). *Interpreting as interaction*. London and New York: Longman.

Wadensjö, C. (2014). *Interpreting as interaction*. Routledge. Book.

Wadhawan, A. (2023). *Simultaneous speech to speech translation* (Doctoral dissertation, University of Pennsylvania). <https://www.cis.upenn.edu/~ccb/publications/masters-theses/Anshul-Wadhawan-masters-thesis-2023.pdf>.

Wang, C. (2017). Professional interpreters = cultural mediators? *TranslatoLogica*, 1(2017), 93-114. <https://discovery.ucl.ac.uk/id/eprint/10086943/>.

Wang, C., Inaguma, H., Chen, P. J., Kulikov, I., Tang, Y., Hsu, W. N., ... & Pino, J. (2022). Simple and effective unsupervised speech translation. *arXiv preprint arXiv:2210.10191*. <https://arxiv.org/abs/2210.10191>

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. *Engineering*, 18, 143-153. <https://www.sciencedirect.com/science/article/pii/S2095809921002745>

Wang, Y., Bai, J., Huang, R., Li, R., Hong, Z., & Zhao, Z. (2023). Speech-to-speech translation with discrete-unit-based style transfer. *arXiv preprint arXiv:2309.07566*. <https://arxiv.org/abs/2309.07566>

Wang, Y., Stanton, D., Zhang, Y., Ryan, R. S., Battenberg, E., Shor, J., ... & Saurous, R. A. (2018, July). Style tokens: Unsupervised style modeling, control and transfer in end-to-end speech synthesis. In *International conference on machine learning* (pp. 5180-5189). PMLR. <https://proceedings.mlr.press/v80/wang18h.html?ref=https://githubhelp.com>

Weinstein, C. J. (2002). *Speech-to-speech translation: Technology and applications study* (Technical Report No. 1080). MIT Lincoln Laboratory. <https://apps.dtic.mil/sti/pdfs/ADA401684.pdf>

Weiss, R. J., Skerry-Ryan, R. J., Battenberg, E., Mariooryad, S., & Kingma, D. P. (2021, June). Wave-tacotron: Spectrogram-free end-to-end text-to-speech synthesis. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5679-5683). IEEE. http://bengio.abracadoudou.com/cv/publications/pdf/wang_2017_arxiv.pdf

Welleck, S., Weston, J., Szlam, A., & Cho, K. (2019, July). Dialogue natural language inference. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3731-3741). <https://aclanthology.org/P19-1363/>

Wester, M., Dines, J., Gibson, M., Liang, H., Wu, Y. J., Saheer, L., ... & Yamagishi, J. (2010). Speaker adaptation and the evaluation of speaker similarity in the EMIME speech-to-speech translation project. In *Proc. of 7th ISCA Speech Synthesis Workshop*. <https://www.research.ed.ac.uk/en/publications/speaker-adaptation-and-the-evaluation-of-speaker-similarity-in-th>

Wester, M., Kessens, J. M., & Strik, H. (2000). Pronunciation variation in ASR: Which variation to model?. <https://repository.uhn.ru.nl/bitstream/handle/2066/76193/76193.pdf>

Williams, J., Pizzi, K., Das, S., & Noé, P. G. (2023). New challenges for content privacy in speech and audio. *arXiv preprint arXiv:2301.08925*. <https://arxiv.org/abs/2301.08925>

Wilss, W. (1999). *Translation and Interpreting in the 20th Century*. <https://www.torrossa.com/gs/resourceProxy?an=5016202&publisher=FZ4850>.

WRHA. (2021, September). Code of Ethics and Standards of Practice for Professional interpreters. professionals.wrha.mb.ca/old/professionals/language/files/EthicsCode.pdf

Wu, H., Chen, X., Lin, Y. C., Chang, K. W., Chung, H. L., Liu, A. H., & Lee, H. Y. (2024). Towards audio language modeling--an overview. *arXiv preprint arXiv:2402.13236*. <https://arxiv.org/abs/2402.13236>.

Xiao, W., Huang, Z., Gan, L., He, W., Li, H., Yu, Z., ... & Zhu, L. (2025, April). Detecting and mitigating hallucination in large vision language models via fine-grained ai feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 24, pp. 25543-25551). <https://doi.org/10.1609/aaai.v39i24.34744>

Xiong, H., He, Z., Wu, H., & Wang, H. (2019, July). Modeling coherence for discourse neural machine translation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 7338-7345). <https://ojs.aaai.org/index.php/AAAI/article/view/4721>

Xu, R. (2018). Corpus-based terminological preparation for simultaneous interpreting. *Interpreting*, 20(1), 29-58. <https://www.jbe-platform.com/content/journals/10.1075/intp.00002.xu>.

Xu, R. (2015). *Terminology preparation for simultaneous professional interpreters* (Doctoral dissertation, University of Leeds). <https://etheses.whiterose.ac.uk/id/eprint/10164/>.

Xuan, G., Zhang, W., & Chai, P. (2001, October). EM algorithms of Gaussian mixture model and hidden Markov model. In *Proceedings 2001 international conference on image processing (Cat. No. 01CH37205)* (Vol. 1, pp. 145-148). IEEE. <https://ieeexplore.ieee.org/abstract/document/958974/>

Yan, L., Echeverria, V., Fernandez-Nieto, G. M., Jin, Y., Swiecki, Z., Zhao, L., ... & Martinez-Maldonado, R. (2024, May). Human-AI collaboration in thematic analysis using ChatGPT: A user study and design recommendations. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-7). <https://dl.acm.org/doi/abs/10.1145/3613905.3650732>.

Yang, X., Zhan, R., Wong, D. F., Wu, J., & Chao, L. S. (2023). Human-in-the-loop machine translation with large language model. *arXiv preprint arXiv:2310.08908*. <https://arxiv.org/abs/2310.08908>

Yao, B., Jiang, M., Bobinac, T., Yang, D., & Hu, J. (2024, November). Benchmarking machine translation with cultural awareness. In *Findings of the Association for Computational Linguistics: EMNLP 2024* (pp. 13078-13096). <https://aclanthology.org/2024.findings-emnlp.765/>

Yi, R. (2024). Manner matters: linguistic equity through a court professional interpreter in Australia. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 1-23. <https://link.springer.com/article/10.1007/s11196-023-10090-3>.

Yule, G. (2022) *The study of language*. Cambridge university press. https://assets.cambridge.org/97811070/44197/excerpt/9781107044197_excerpt.pdf.

Yun, S., Lee, Y. J., & Kim, S. H. (2014). Multilingual speech-to-speech translation system for mobile consumer devices. *IEEE Transactions on Consumer Electronics*, 60(3), 508-516. <https://ieeexplore.ieee.org/abstract/document/6937337/>

Zampieri, M., Nakov, P., & Scherrer, Y. (2020). Natural language processing for similar languages, varieties, and dialects: A survey. *Natural Language Engineering*, 26(6), 595-612. <https://www.cambridge.org/core/journals/natural-language-engineering/article/natural-language->

[processing-for-similar-languages-varieties-and-dialects-a-survey/229652C86E329F83346BB6C66B9521A6](https://www.jbe-platform.com/content/journals/10.1075/intp.14.2.04zha)

Zhan, C. (2012). Mediation through personal pronoun shifts in dialogue interpreting of political meetings. *Interpreting*, 14(2), 192-216. <https://www.jbe-platform.com/content/journals/10.1075/intp.14.2.04zha>

Zhang, H., Sproat, R., Ng, A. H., Stahlberg, F., Peng, X., Gorman, K., & Roark, B. (2019). Neural models of text normalization for speech applications. *Computational Linguistics*, 45(2), 293-337. <https://direct.mit.edu/coli/article-abstract/45/2/293/1637>.

Zhang, R., & Liu, J. (2024). Effects and Causes of VR-Supported Interpreting Learning Environment on the Interpretation Classroom Anxiety of Student Professional interpreter. *Open Journal of Applied Sciences*, 14(2), 398-410. <https://www.scirp.org/journal/paperinformation?paperid=131345>.

Zhang, S., Mehrish, A., Li, Y., & Poria, S. (2025). PROEMO: Prompt-Driven Text-to-Speech Synthesis Based on Emotion and Intensity Control. *arXiv preprint arXiv:2501.06276*. <https://arxiv.org/abs/2501.06276>

Zheng, J., Hong, H., Liu, F., Wang, X., Su, J., Liang, Y., & Wu, S. (2024). Fine-tuning large language models for domain-specific machine translation. *arXiv preprint arXiv:2402.15061*. <https://arxiv.org/abs/2402.15061>

Zhou, N., Bamman, D., & Bleaman, I. L. (2025). Culture is not trivia: Sociocultural theory for cultural nlp. *arXiv preprint arXiv:2502.12057*.

Zhou, R., Ito, A., & Nose, T. (2024). Preserving Speaker Information in Direct Speech-to-Speech Translation with Non-Autoregressive Generation and Pretraining. *arXiv preprint arXiv:2412.07316*. <https://arxiv.org/abs/2412.07316>

Zielnik-Kołodzińska, R. (2021). Linguistic Mediation in the Past and in Future. *Rödl & Partner*. <https://www.roedl.com/insights/professional-translation-services-linguistic-mediation?> Retrieve 23 March 2025.

Zinukova, N. (2016). Acquiring translation and interpreting competences: integrated approach. *Іноземні мови*, (3), 18-24. http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/im_2016_3_4.pdf.

Zoph, B., Yuret, D., May, J., & Knight, K. (2016). Transfer learning for low-resource neural machine translation. *arXiv preprint arXiv:1604.02201*. <https://arxiv.org/abs/1604.02201>

Zuluaga-Gomez, J., Prasad, A., Nigmatulina, I., Sarfjoo, S. S., Motlicek, P., Kleinert, M., ... & Zhan, Q. (2023, January). How does pre-trained wav2vec 2.0 perform on domain-shifted asr? an extensive benchmark on air traffic control communications. In *2022 IEEE spoken language technology workshop (SLT)* (pp. 205-212). IEEE. <https://ieeexplore.ieee.org/abstract/document/10022724/>

APPENDIX I - Script

To run the download automation script, it is first necessary to install Python programming language on the computer (van Rossum & Drake, 2009). Then, the libraries “requests” and “serpAPI” must be installed, which can be done giving the prompt “pip install requests serpapi” (Saabith et al., 2021).

These libraries are installed directly through the computer’s command terminal, which, in the Windows operating system, can be accessed by pressing “Win + R + cmd + enter”. The “requests” library is a Python module for making HTTP requests. It allows sending requests to web servers and retrieving responses, such as HTML pages, JSON files, and even PDFs (Kumar et al., 2022).

The “serpAPI” library searches for articles on GS, retrieves news, images, e-commerce data, maps, and other Google-related information. It automates searches and extracts data in a structured manner (Jay, 2024).

Once the libraries are installed, the next step is to create an account on SerpAPI and obtain the API key, which is an alphanumeric code required to search on GS.

The free code editing tool Jupyter Notebook (Dombrowski et al., 2023) is used to execute the download automation script. It allows users to write and run Python code directly in cells, similar to an interactive notebook. This tool is used in data science, machine learning, and automation, as it enables working with code, graphs, and annotations in the same document.

Table 21 - Download automation of the academic publications on S2ST.

```
import requests
import os
from serpapi import GoogleSearch

def search_scholar(doi, api_key):
    """
    Retrieves articles from Google Scholar via SerpAPI using
    a DOI.
    """
    params = {
        "engine": "google_scholar",
        "q": doi,
        "api_key": api_key
    }

    search = GoogleSearch(params)
```

```

results = search.get_dict()

if "organic_results" in results:
    for result in results["organic_results"]:
        if "link" in result and ".pdf" in
            result["link"]:
                return result["link"]

return None

def download_pdf(pdf_url, save_path):
    """
    Retrieves the article's PDF and saves it to a local
    directory.
    """
    try:
        response = requests.get(pdf_url, stream=True)
        if response.status_code == 200:
            with open(save_path, "wb") as file:
                for chunk in response.iter_content(1024):
                    file.write(chunk)
            print(f"Download concluded: {save_path}")
        else:
            print("Error")
    except Exception as e:
        print(f"Error: {e}")

def main():
    api_key = "YOUR_SERPAPI_KEY" # Replace with your API key
    dois = [
        "10.1038/s41586-020-2649-2", # DOI Example
        "10.1126/science.abd7728"
    ]

    save_directory = "downloads"
    os.makedirs(save_directory, exist_ok=True)

    for doi in dois:
        print(f"Searching paper for DOI: {doi}")
        pdf_url = search_scholar(doi, api_key)

        if pdf_url:
            filename = f"{doi.replace('/', '_')}.pdf"
            save_path = os.path.join(save_directory,
                                      filename)
            download_pdf(pdf_url, save_path)
        else:
            print(f"No PDF was found to {doi}")

if __name__ == "__main__":
    main()

```

Source: the author.

APPENDIX II - Literature Review Generated by ChatGPT

Model: 4o

Words: 1,352

1. Historical Evolution of the Professional interpreter's Role

The evolution of the professional interpreter's role is closely linked to historical needs such as trade, diplomacy, colonization, and international relations. From its origins as ad hoc language broker, the role has transformed into highly trained professionals working within institutions. This shift is also connected to the recognition of interpreting as a profession and academic field.

Ancient and pre-modern eras

Professional interpreters in ancient civilizations, such as Ancient Egypt, held significant positions. Records show that professional interpreters, referred to as “dragomans,” were recognized officials, even having their own hierarchy and oversight.

Genesis 42 shows professional interpreters were used in ancient governance and trade, as Joseph used one when speaking to his Hebrew brothers in Egypt. In ancient Greece and Rome, professional interpreters were essential but not always trusted (hence the Latin proverb “*Interpres nullus amicus*”, “no professional interpreter is a friend,” reflecting suspicion of their loyalty).

In the Middle Ages, professional interpreters were essential for travel, missionary work, and exploration. During the Age of Discovery, figures like La Malinche and Sacagawea played key roles as cultural and linguistic intermediaries. These professional interpreters often learned languages informally and operated under significant power imbalances, sometimes revered, sometimes reviled. Their role was not yet a profession, but a situational function performed with little training.

The birth of conference interpreting

Before World War I, French was commonly used in international diplomacy. However, after the 1919 Paris Peace Conference, English was also introduced, leading to the use of consecutive professional interpreters for high-level negotiations.

The League of Nations in the 1920s used professional interpreters, who were pioneers in developing interpreting techniques but interpreting was not yet a recognized profession.

The Nuremberg Trials (1945-46) marked a turning point in interpreting. To conduct trials efficiently in four languages, simultaneous interpreting technology was introduced using headsets, microphones, and IBM equipment. This innovation drastically reduced trial time and proved interpreting could keep pace with live speech. Simultaneous interpretation was pioneered at Nuremberg and adopted by the UN, and demand for professional interpreters grew during the Cold War, decolonization, and expansion of organizations like the European Common Market.

Professionalization and training

The International Association of Conference Interpreters (AIIC), founded in 1953, formalized the profession by establishing a code of ethics, working conditions, and recruitment standards. Specialized professional interpreter training schools like ESIT in Paris developed curricula focusing on interpreting techniques, domain knowledge, and professional ethics. Graduates from these programs became the new generation of professional interpreters staffing international institutions.

In the mid-to-late 20th century, professional interpreters were seen as skilled language specialists who should be unobtrusive. Seleskovitch's theory in the 1960s-70s emphasized understanding meaning but still viewed the professional interpreter as a non-participating narrator. In the West, professional interpreters gained prestige and were sometimes glamorized.

Court and community interpreting lacked professional development for years. In the US, court interpreting was unaddressed until the 1970s, when the needs of non-English speakers led to the Federal Court Professional interpreter Act and certification programs. Community interpreting was similarly informal, with untrained volunteers often used. Growing awareness of immigrants' rights led to investment in training and codes of practice for community professional interpreters by the 1990s.

Interpreting Studies as a field

Early works on interpreting (1950s–70s) often portrayed professional interpreters as neutral messengers. In the 1990s, researchers began to view professional interpreters not as invisible conduits, but as active participants in shaping communication.

The "conduit model" of professional interpreters as passive messengers has been replaced by a view that acknowledges their agency and negotiated role, as evidenced by historical transcripts revealing their influence on outcomes.

Professional interpreters' status has risen from being viewed as "necessary evils" to "essential professionals." The institutionalization of professional interpreters in courts and public services in many countries today reflects this evolution. Interpreting is a mature profession with professional organizations, career paths, and academic research.

In the past, professional interpreters were informal, all-purpose cultural intermediaries; now, they are systematic professionals. Professional interpreters today are specialized, ethical, and well-trained professionals who actively coordinate communication.

2. Social, Ethical, and Professional Identity of Professional interpreters

Professional interpreters do not work in a vacuum; they are social actors embedded in communicative events that often involve disparities in power, culture, and expectations. The professional interpreter's social role, ethics, and professional identity are key areas of debate in Interpreting Studies. These debates focus on the professional interpreter's level of participation and how they balance their responsibilities.

Neutrality vs. advocacy

The traditional view of professional interpreters, especially in conference and legal settings, was that they should be neutral and uninvolved in the interaction, simply transmitting information. Codes of ethics require impartiality; professional interpreters must not give personal opinions or advice.

While professional interpreters are ideally unobtrusive, research shows they inevitably influence communication and that all parties are somewhat aware of this. Professional

interpreters often play an active role in facilitating communication and building trust, rather than being invisible translators.

The concept of alignment is controversial in interpreting. While codes demand neutrality, professional interpreters may sympathize with one side, subtly influencing their renditions. For example, a professional interpreter might soften a doctor's harsh words to a patient out of compassion. Some scholars argue that strict neutrality can be unethical if it leads to harm, and suggest advocacy to ensure the marginalized voice is heard.

Professional interpreters occupy a third space, neither neutral nor advocate, but a mediator with an active decision-making role. Jacobsen (2003) found that legal professional interpreters prioritize conveying intended meaning, even if it requires deviations from the source text, thus contradicting the pure conduit ideal.

Ethical frameworks:

Professional ethics for professional interpreters include accuracy, confidentiality, impartiality, and professional demeanor, often outlined in codes by organizations like AIIC, RID, and NAJIT.

Ethical principles for professional interpreters include confidentiality, impartiality, and accuracy. However, applying these principles can be challenging. For example, is it better to clarify a confusing term (deviating from verbatim accuracy) or stick to the words (risking misunderstanding)? Different codes give different guidance, adding to the complexity.

Professional interpreters must also manage role boundaries and avoid pressure from others to step out of their role. Professional ethics are crucial for professional interpreters to manage role boundary issues and maintain impartiality, especially in community and medical settings where they might be treated as co-interviewers or aides.

The concept of "role-space" describes the professional interpreter's role as a dynamic and multi-dimensional space, influenced by factors like self-presentation, involvement, and alignment with participants. The professional interpreter's role can vary depending on the situation. For example, in a simple bank transaction, the professional interpreter may only need to translate directly.

However, in a sensitive counseling session, the professional interpreter may need to mediate to ensure respect and understanding. This shows that professional interpreters have professional autonomy to make decisions based on the situation.

Professional identity

As interpreting has become more professionalized, professional interpreters have shifted their self-perception and projection from informal helpers to skilled experts. Professionalization of interpreting has led to a shared identity among professional interpreters.

Pöchhacker notes that professional interpreters in the late 20th century became more specialized and professionalized, leading to distinct career paths and sometimes tension between groups (e.g., conference vs. community professional interpreters). However, there is a growing sense of unity, especially as professional interpreters face common issues.

Professional interpreters develop their identity through training and socialization, where they are taught specific role philosophies. A community interpreting program may promote the idea of the professional interpreter as a cultural mediator, shaping students' professional identities. These identities can evolve with experience, as professional interpreters may adapt their approach over time.

Situational and social context

A professional interpreter's social identity can affect their role performance and may lead to role conflict. Sign language professional interpreters' roles have evolved from helper (pre-1970s) to communication facilitator (post-professionalization) to bicultural mediator (current), recognizing the need to bridge Deaf and hearing cultures.

Deaf community norms influence sign language professional interpreters' ethos, pushing them beyond neutrality towards cultural sensitivity and full Deaf participation. Sign language professional interpreters may quickly align with and advocate for Deaf clients' needs, which is accepted in the field.

Professional interpreters must balance ethical standards with the unique human and contextual factors of each interpreted event. A modern professional interpreter is a reflective practitioner who is aware of their role, makes ethical choices, and continuously improves their practice. Professional interpreters are evolving from language conduits to language mediators,

signifying a shift towards acknowledging their agency and responsibility in shaping communication's social dynamics.

Sources:

- Pöchhacker, F. (2022). *Introducing Interpreting Studies*. London: Routledge.
- Mikkelsen, H. (2008). "Evolving Views of the Court Professional interpreter's Role: Between Scylla and Charybdis." *Beyond the Ivory Tower: Rethinking Translation Pedagogy*.
- Angelelli, C. (2004). *Revisiting the Professional interpreter's Role*. Amsterdam: John Benjamins.
- Napier, J., et al. (2015). *The Routledge Handbook of Interpreting*. New York: Routledge. ([Here or There | Gallaudet University Press](#))
- Wadensjö, C. (1998). *Interpreting as Interaction*. London: Longman.
- Gile, D. (1995). *Basic Concepts and Models for Professional interpreter and Translator Training*. Amsterdam: John Benjamins. ([Gile's Effort Models for interpreting](#)) ([Gile's Effort Models for interpreting](#))
- Jacobsen, B. (2003). "Pragmatic Meaning in Court Interpreting." *Perspectives*, 11(3).
- Baigorri-Jalón, J. (2014). *From Paris to Nuremberg: The Birth of Conference Interpreting*. ([The Nuremberg Trials & Simultaneous Interpretation | Eriksen Translations](#))
- Franz Pöchhacker (2022). "Professional interpreters and interpreting: shifting the balance?" *The Translator*, 28(2). ([\(PDF\) Interpreting technologized: Distance and assistance](#)) ([\(PDF\) Interpreting technologized: Distance and assistance](#))
- Ingrid Piller (2021). "The interpreting profession in Ancient Egypt." *Language on the Move* blog. ([The interpreting profession in Ancient Egypt – Language on the Move](#))