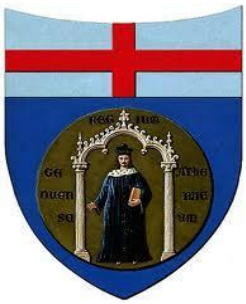
UNIVERSITÀ DI GENOVA



Ph.D. in Digital Humanities

Languages, Cultures and Digital Technologies, Cycle XXXIV

Department of Modern Languages and Cultures

Doctoral Thesis

'Testing quality in interlingual respeaking and other methods of interlingual live subtitling'

dott.ssa Alice Pagano

2022

Supervisor:

prof.ssa Anna Giaufret

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AD	Audio description
AS	Accessibility Studies
ASR	Automatic Speech Recognition
AT	Automatic Translation
AV	Audio Visual
AVT	Audiovisual Translation
B/VIP	Blind and Visually Impaired
CAI	Computer-assisted Interpreting
CAT	Computer-assisted Translation
CNN	Convolutional Neural Network
Cont-omiss	Content omission
Cont-add	Content addition
Cont-subs	Content substitution
DHOH	Deaf and hard of hearing
DL	Deep Learning
EE	Effective Edition
ECTS	European Credit Transfer and Accumulation System
EX1	First experiment
EX2	Second experiment
Form-corr	Form-correctness
GNMT	Google Neural Machine Translation
HCI	Human-Computer Interaction
HE	Higher Education
HMI	Human-Machine Interaction
ILS	Interlingual Live Subtitling

ILSA	Interlingual Live Subtitling for Access
INT	Interpreter
IT	Information Technology
LIS	Italian Sign Language
LTA	Live Text Access
MA	Media Accessibility
MOOC	Massive Open Online Course
MT	Machine Translation
NMT	Neural Machine Translation
NDACA	National Disability Arts Collection and Archive
NLU	Natural Language Understanding
PBSMT	Phrase-Based Statistical Machine Translation
RBSMT	Rule-Based Statistical Machine Translation
RESP	Respeaker
RNN	Recurrent Neural Network
RQ	Research Question
SDH	Subtitling for the deaf and hard of hearing
SI	Simultaneous interpreting
SL	Source language
SLS	Same-language subtitling
SMART	Shaping Multilingual Access through Respeaking Technology
SMT	Statistical Machine Translation
SR	Speech Recognition
ST	Source text
TL	Target language
TS	Translation Studies
TT	Target text
UNCRPD	United Nations Convention on the Rights of Persons with Disabilities
WPM	Words per minute

List of Appendices[1]

Appendix 1 – First experiment materials (NTR analyses, video and audio recordings)

Appendix 2 – Pre-workshop questionnaire for the first experiment

Appendix 3 – Post-workshop and pre-experiment questionnaire for the first experiment

Appendix 4 – Post-experiment questionnaire for the first experiment

Appendix 5 – Second experiment materials (NTR analyses, video and audio recordings)

Appendix 6 – Pre-workshop questionnaire for the second experiment

Appendix 7 – Post-workshop and pre-experiment questionnaire for the second experiment

Appendix 8 – Post-experiment questionnaire for the second experiment

Appendix 9 – Audio and video recordings authorization form

[1] The appendices referring to questionnaires are available at the end of the thesis translated in English. Original questionnaires in Italian and the participants' answers are available in the reference appendix at a separate link. Some answers are partially translated in English by the researcher in Sections 3.3 and 4.4 in Chapter 3.

Abstract

Live subtitling (LS) finds its foundations in pre-recorded subtitling for the d/Deaf and hard of hearing (SDH) to produce real-time subtitles for live events and programs. LS implies the transfer from oral into written content (intersemiotic translation) and can be carried out from and to the same language (intralingual), or from one language to another (interlingual) to provide full accessibility for all, therefore combining SDH to the need of guaranteeing multilingual access as well. Interlingual Live Subtitling (from now on referred to as ILS) in real-time is currently being achieved by using different methods: the focus here is placed on interlingual respeaking as one of the currently used methods of LS – also referred to in this work as speech-to-text interpreting (STTI) – which has triggered growing interest also in the Italian industry over the past years.

The hereby presented doctoral thesis intends to provide a wider picture of the literature and the research on intralingual and interlingual respeaking to the date, emphasizing the current situation in Italy in this practice.

The aim of the research was to explore different ILS methods through their strengths and weaknesses, in an attempt to inform the industry on the impact that both potentialities and risks can have on the final overall quality of the subtitles with the involvement of different techniques in producing ILS. To do so, five ILS workflows requiring human and machine interaction to different extents were tested overall in terms of quality, thus not only from a linguistic accuracy point of view, but also considering another crucial factor such as delay in the broadcast of the subtitles. Two case studies were carried out with different language pairs: a first experiment (English to Italian) tested and assessed quality in interlingual respeaking on one hand, then simultaneous interpreting (SI) combined with intralingual respeaking, and SI and Automatic Speech Recognition (ASR) on the other. A second experiment (Spanish to Italian) evaluated and compared all the five methods: the first three

again, and two others more machine-centered: intralingual respeaking combined with machine translation (MT), and ASR with MT.
Two workshops in interlingual respeaking were offered at the master's degree in Translation and Interpreting from the University of Genova to prepare students for the experiments, aimed at testing different training modules on ILS and their effectiveness on students' learning outcomes. For the final experiments, students were assigned different roles for each tested method and performed different required tasks producing ILS from the same source text: a video of a full original speech at a live event. The obtained outputs were analyzed using the NTR model (Romero-Fresco & Pöchhacker, 2017) and the delay was calculated for each method.

Preliminary quantitative results deriving from the NTR analyses and the calculation of delay were compared to other two case studies conducted by the University of Vigo and the University of Surrey, showing that more and fully-automated workflows are, indeed, faster than the others, while they still present several important issues in translation and punctuation. Albeit on a small scale, the research also shows how urgent and potentially easy could be to educate translators and interpreters in respeaking during their training phase, given their keen interest in the subject matter.

It is hoped that the results obtained can better shed light on the repercussions of the use of different methods and induce further reflection on the importance of human interaction with automatic machine systems in providing high quality accessibility at live events. It is also hoped that involved students' interest in this field, which was completely unknown to them prior to this research, can inform on the urgency of raising students' awareness and competence acquisition in the field of live subtitling through respeaking.

Keywords: Accessibility; Live subtitling; Speech-to-text interpreting; Intralingual Respeaking; Interlingual Respeaking; simultaneous interpreting; Automatic Speech Recognition; Machine Translation.

Riassunto

La sottotitolazione in tempo reale (*Live Subtitling*, LS), trova le sue fondamenta nella sottotitolazione preregistrata per non udenti e ipoudenti per la produzione di sottotitoli per eventi o programmi televisivi dal vivo. La sottotitolazione live comporta il trasferimento da un contenuto orale a uno scritto (traduzione intersemiotica) e può essere effettuata da e verso la stessa lingua (intralinguistica), o da una lingua a un'altra (interlinguistica), fornendo così accessibilità per soggetti non udenti e al tempo stesso garantendo accesso multilingue ai contenuti audiovisivi. La sottotitolazione interlinguistica in tempo reale (d'ora in poi indicata come ILS, *Interlingual Live Subtitling*) viene attualmente realizzata con diversi metodi: l'attenzione è qui posta sulla tecnica del respeaking interlinguistico, uno dei metodi di sottotitolazione in tempo reale o *speech-to-text interpreting* (STTI) che ha suscitato negli ultimi anni un crescente interesse, anche nel panorama italiano.

Questa tesi di Dottorato intende fornire un quadro della letteratura e della ricerca sul respeaking intralinguistico e interlinguistico fino ad oggi, con particolare enfasi sulla situazione attuale in Italia di questa pratica.
L'obiettivo della ricerca è stato quello di esplorare diversi metodi di ILS, mettendone in luce i punti di forza e le debolezze nel tentativo di informare il settore delle potenzialità e dei rischi che possono riflettersi sulla qualità complessiva finale dei sottotitoli attraverso l'utilizzo di diverse tecniche. Per fare ciò, sono stati testati in totale cinque metodi di ILS con diversi gradi di interazione uomo-macchina; ciascun metodo è stato analizzato in termini di qualità, quindi non solo dal punto di vista dell'accuratezza linguistica, ma anche considerando un altro fattore cruciale quale il ritardo nella trasmissione dei sottotitoli stessi. Nello svolgimento della ricerca sono stati condotti due casi di studio con diverse coppie linguistiche: il primo esperimento (dall'inglese all'italiano) ha testato e valutato la qualità di respeaking interlinguistico, interpretazione simultanea insieme a respeaking intralinguistico

e, infine, interpretazione simultanea e sistema di riconoscimento automatico del parlato (*Automatic Speech Recognition*, ASR). Il secondo esperimento (dallo spagnolo all'italiano) ha valutato e confrontato cinque i metodi: i primi tre appena menzionati e altri due in cui la macchina svolgeva la maggior parte se non la totalità del lavoro: respeaking intralinguistico e traduzione automatica (*Machine Translation*, MT), e ASR con MT.

Sono stati offerti due laboratori di respeaking interlinguistico nel Corso magistrale in Traduzione e Interpretazione dell'Università di Genova per preparare gli studenti agli esperimenti, volti a testare diversi moduli di formazione sull'ILS e la loro efficacia sull'apprendimento degli studenti. Durante le fasi di test, agli studenti sono stati assegnati diversi ruoli per ogni metodo, producendo sottotitoli interlinguistici live a partire dallo stesso testo di partenza: un video di un discorso originale completo durante un evento dal vivo. Le trascrizioni ottenute, sotto forma di sottotitoli, sono state analizzate utilizzando il modello NTR (Romero-Fresco & Pöchhacker, 2017) e per ciascun metodo è anche stato calcolato il ritardo.

I risultati quantitativi preliminari derivanti dalle analisi NTR e dal calcolo del ritardo sono stati confrontati con altri due casi di studio condotti dall'Università di Vigo (Spagna) e dall'Università del Surrey (Gran Bretagna), sottolineando come i flussi di lavoro più automatizzati o completamente automatizzati siano effettivamente più veloci degli altri, ma al contempo presentino ancora diversi problemi di traduzione e di punteggiatura. Anche se su scala ridotta, la ricerca dimostra anche quanto sia urgente e possa potenzialmente essere facile formare i traduttori e gli interpreti sul respeaking durante il loro percorso accademico, grazie anche al loro spiccato interesse per la materia.
Si spera che i risultati ottenuti possano meglio mettere in luce le ripercussioni dell'uso dei diversi metodi a confronto, nonché indurre un'ulteriore riflessione sull'importanza dell'interazione umana con i sistemi automatici di traduzione e di riconoscimento del parlato nel fornire accessibilità di alta qualità per eventi dal vivo. Si spera inoltre che l'interesse degli studenti in questo campo, che era a loro completamente sconosciuto prima di questa ricerca, possa informare sull'urgenza di sensibilizzare gli studenti nel campo della sottotitolazione dal vivo attraverso il respeaking.

Parole chiave: accessibilità; sottotitolazione in tempo reale; interpretazione *speech-to-text*; respeaking intralinguistico; respeaking interlinguistico; interpretazione simultanea; riconoscimento automatico del parlato; traduzione automatica.

Introduction

1. Interlingual live subtitles and accessibility

When considering accessibility in its broader sense, including all types of inabilities or limited abilities in performing physical, sensorial, social or cognitive functions, an important insight is given by the perspective through which we perceive accessibility itself. The interconnection between accessibility and disability (or impairment, or limitation) implies the former being provided when the latter is present, according to different perspectives. In this regard, two main contrasting approaches to the matter were conceptualized over time as models, to which specific attention is dedicated in the following lines: the medical model and the social model.

With the social and the medical approaches in the spotlight, an interesting video by the NDACA (National Disability Arts Collection & Archive)[2] depicts how, according to the Medical Model of disability, the word 'disabled', meaning 'less able', identifies the person with an impairment as responsible for overcoming these disabling barriers, and "[i]t regards the limitations faced by people with disabilities as resulting primarily, or solely, from their impairments" (Wasserman *et al.*, 2022). The later proposed Social Model of disability, though, shifts the perspective between the individual and their social environment, so that people do not 'have a disability', they are instead 'disabled by society'. In other words, it is the physical and social barriers created by society when not recognizing the need for accessibility that hinder disabled people, not their nature. It is more, it is the way accessibility is (or is not) provided that allows (or does not allow) them to properly access what other people can.

[2] This video was retrieved from the ILSA course, Module 1, at the following link: http://ka2-ilsa.webs.uvigo.es/course-and-training-materials/.

Over the years different models of disability were developed apart from the medical and the social ones, such as the older moral and religious model, the cultural model (differing from the social and the medical ones as not focusing on only one factor but rather different notions of disability in a specific culture), the identity and the human rights models (with close affinity to the social model), or the charity model, in which disabled people are pitied and victimized (Retief & Letsosa, 2018).

Two main strands of the social model can be identified: the minority group model, which sees people with disabilities as a minority subject to stigmatization and exclusion, and the human variation model, according to which many of the challenges faced by disabled people do not depend on their exclusion, but from a disparity between their characteristics and the physical and social environment.
More moderate views of the social model, anyways, tend to treat disability as an interaction between both biological and social causes despite the most frequent, and not exclusively as societal obstacle for disabled people. As Wasserman (2022) explains, the social model in all its more or less extreme positions can be mainly criticized since it does not cater for a proper classification of disability, which is instead needed to distinguish disability discrimination from other types of discriminations, and that completely overlooking impairment as an objective basis for classification makes no sense.
Over the years, accessibility has acquired a universalist instead of a particularist focus, as well as a user-centered instead of a maker-centered approach (Greco, 2018), thereby changing perspectives and making accessibility a concern of the many, not the few, and consequently of society as a whole. This is also thanks to many resolutions that have been adopted to meet the goal, such as the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) in 2006, the EU Audiovisual Media Service Directive in 2010, amended in 2018, and the European Accessibility Act which entered into force in 2019.

In perceiving disability as a status for which society needs to enable access in all its forms and declinations as a human right, the focus of this research is on sensory disability, and specifically on hearing disability, namely for the d/Deaf and Hard of Hearing (DHOH),

people with hearing loss and/or impairment who cannot access audio contents or can but only to a certain extent.

As we move into the field of Audiovisual Translation (AVT) which has witnessed a surprisingly growing demand in the past twenty years,

The focus is placed on the specific branch referred to as Media Accessibility (MA) in reference to screen translation strictly connected with the need for accessibility, something that acquires even more relevance in our interconnected world with fast-changing communicative needs, and in this ever-connected society that, especially over the last two pandemic years, has called even more for a barrier free cross-cultural communication (Romero-Fresco & Alonso-Bacigalupe, 2022).

This research deals with Media Accessibility (MA) for accessible multilingual communication, covering accessibility through the creation of interlingual subtitles for live events – i.e., from audio input into written output, and from one language into another. This way, access is provided for audiences with and also without hearing disability (Romero-Fresco, 2018): MA accounts, indeed, not only for sensory – such as subtitling for the d/Deaf and hard of hearing (SDH) – but also to linguistic barriers (Díaz Cintas, 2005), where the cross-cultural factors need to be addressed in a similar manner to simultaneous interpreting (SI). Interlingual live subtitling (ILS) does provide accessibility in such terms – as SDH does, with a written product – also offering a multilingual access to the content in real-time. Figure 1 shows the shared tasks of the three disciplines – interpreting, live and pre-recorded subtitling – intertwining in producing ILS.

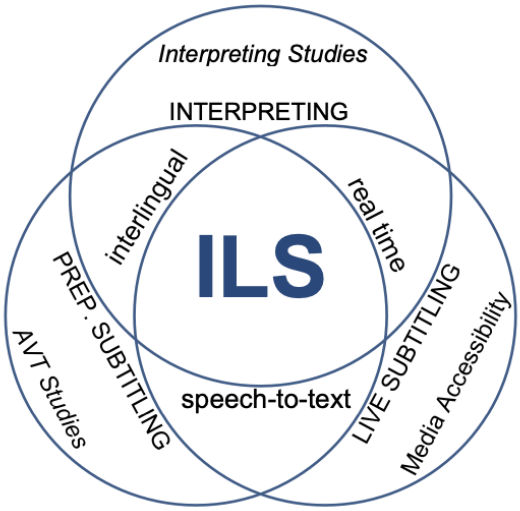


Figure 1 – Interlingual Live Subtitling at the crossroad of other disciplines
(Pöchhacker & Remael, 2019: 132)

In the past two decades audio contents have been made more and more accessible not only in official and institutional contexts, but also through public and private TV programs (Eugeni & Bernabé, 2019) and the demand for subtitles produced live is growing. Among different techniques, one of the most recent and relatively new to produce live subtitles (LS) and that is paving the way both for intralingual and interlingual contexts is respeaking. Situated within AVT and MA (Dawson, 2020) and at the crossroad between pre-recorded subtitling and SI, respeaking is defined as:

> "a technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the Deaf and Hard of Hearing audience, to a speech recognition software, which turns the recognized utterances into subtitles displayed on screen with the shortest possible delay".
>
> (Romero-Fresco, 2011: 1)

Respeaking is a technology-enabled hybrid modality of translation (Davitti & Sandrelli, 2020), that shares a common ground with SI in terms of skills and competences required (Russello, 2013). It does share common ground with subtitling as well. Respeaking was in fact initially carried out only by subtitlers, before interpreters (Romero-Fresco & Eugeni, 2020). The technique is carried out through speech-recognition (SR) software that enables the production of written output.

In our interconnected reality, human capabilities are always supported by Information Technologies (ITs) to great extent and the broader use we make of technology in general is undeniable. In the field of translation and interpreting as well, ITs contribute to its evolution and fusion with other fields, like MA itself (Romero-Fresco & Alonso-Bacigalupe, 2022). The large contribution of machines to translation does simplify and optimize workflows substantially: CAT (Computer-Assisted Translation) tools and – though newer – CAI (Computer-Assisted Interpreting) tools are becoming an integrated part of professional translators and interpreters' lives. For this reason, the urgency to further investigate the outcome quality of partially and especially fully automated processes has increased. Translation technologies are developing and improving very fast, therefore it is crucial to be aware of how to use them, to what extent and how to best (co-)work with them. Starting from this assumption, still requiring investigation is to what degree exactly do (or should) humans rely on the machine to carry out their translating tasks, as well as if machines can carry out those same tasks fully autonomously without human contribution (Eugeni, 2019; 2020; Fantinuoli & Prandi, 2021).

2. Purpose of the research

This research work was inspired by and ran in parallel with two other research projects carried out by the University of Vigo (Romero-Fresco & Alonso-Bacigalupe, 2022) and the University of Surrey (Dawson, 2021). The three projects focused on interlingual

respeaking as one of the newest methods of delivering ILS and aimed at comparing the technique with other ILS workflows with different extents of human-machine interaction:

1) Interlingual respeaking
2) Simultaneous interpreting and intralingual respeaking
3) Simultaneous interpreting and Automatic Speech Recognition
4) Intralingual respeaking and Machine Translation (MT)
5) Automatic Speech Recognition and MT

In this thesis, two experiments were carried out during the Ph.D. to test different methods with different factors. The first main experiment of the research involved 15 participants and was carried out from English to Italian, aiming at comparing only the first three methods mentioned above. A second experiment was carried out from Spanish to Italian and was aimed at testing and comparing the five methods. On both occasions, outputs (sets of transcribed subtitles) obtained were analyzed through the NTR model (Romero-Fresco & Pöchhacker, 2017), bearing in mind that in this case the concept of quality depends not only on linguistic accuracy of translated products, but also from other variables such as delay in the creation of the subtitles, and costs. Therefore, the aim of the experiments was not only to calculate accuracy through the model, but to evaluate the outcomes from a more comprehensive perspective for an overall quality assessment.
In carrying out two projects, the ultimate aim was also to compare (at least for the first three methods coinciding between the two tests) the results obtained with two different language pairs, as well as to compare them with the other international experiments related to the same subject matter.

To conduct the experiments, participants were trained in intra and interlingual respeaking. To achieve this, the first ever introduction to training in interlingual respeaking at the University of Genoa was proposed, a first edition for a. y. 2020/21 in English, and a second

in 2021/22 in Spanish. The purpose of the training was to introduce students to the topic, allow them to carry out some practical exercises and familiarize them with the SR software. The participants of the workshop also participated in the experiments that were carried out directly afterwards, therefore training was preparatory for the experiment stage. Only students – thus not professional respeakers – were available for the experiments, but since the aim was to compare different ILS methods rather than test novices in interlingual respeaking, it was chosen to train them and to test them after reaching a basic level of fluency in respeaking.

3. Research questions

The research questions (RQs) derive from the aim of assessing overall quality for each method in producing interlingual live subtitles, closely monitoring more human-oriented methods (Method 1, interlingual respeaking, and Method 2, SI and intralingual respeaking), more machine-centered ones (Method 3, ASR and MT, and Method 4, intralingual respeaking and MT), and, lastly, a fully automated mode (Method 5, ASR and MT).

Respeaking is currently the preferred technique to create LS (Romero-Fresco, 2011), while much research is still to be carried out concerning methods for ILS. Nevertheless, ILS from one language to another through respeaking seems very promising. As described in the following chapters, not only has interlingual respeaking been deemed feasible (Dawson, 2019, 2020; Dawson & Romero-Fresco, 2021), or otherwise promising also among students (Davitti & Sandrelli, 2020), but a research-informed professional profile for interlingual respeakers has been researched and further consolidated in recent years (Pöchhacker & Remael, 2019; Dawson & Romero-Fresco, 2021). As a reflection, a training model for interlingual respeakers has been designed and first created, through two proposals (ILSA project, 2020; Dawson, 2020). ILS through respeaking has good prospects of thriving in the near future thanks to its many similarities with SI, thanks to which training for

interpreters could be improved and implemented with respeaker training to cater to the need for accessibility for a wider audience in live events. More detailed thoughts on the current training situation in ILS and respeaking in Italy and some considerations on interpreters' education for accessibility will be provided in Chapter 1, Section 4.3 and Chapter 5, Section 5.1 of this research.

The main RQ for this thesis is:

How do different ILS methods compare in terms of quality, when quality is referred to as the joint result of both linguistic accuracy and delay in broadcasting, for English and Spanish to Italian language combinations?

To answer the first RQ for this research other RQs can be summarized as follows, according to the two experiments, from now on referred to as EX1 (the first experiment, English to Italian) and EX2 (the second experiment, Spanish to Italian):

EX1 – English to Italian experiment testing three methods (interlingual respeaking; SI + intralingual respeaking; SI + ASR)

RQ1: Among the three considered methods, which one delivers the highest linguistic accuracy?

RQ2: Among the three considered methods, which one is broadcast with minimum delay?

RQ3: Is there, among the three considered methods, one that can provide higher-overall quality ILS, taking into account both accuracy and delay as the main factors for evaluation?

RQ4: Are there other variables to consider in searching for higher-quality methods for ILS, and what are they?

EX2 – Spanish to Italian experiment testing five methods (interlingual respeaking; SI + intralingual respeaking; SI + ASR; intralingual respeaking + MT; ASR + MT)

RQ1: Among the five considered methods, which one delivers the highest linguistic accuracy?

RQ2: Among the five considered methods, which one is broadcast with minimum delay?

RQ3: Is there, among the five considered methods, one that can provide higher-overall quality ILS, taking into account both accuracy and delay as the main factors for evaluation?

RQ4: Are there other variables to consider in searching for higher-quality methods for ILS, and what are they?

Throughout the research, some qualitative data were also collected concerning training in interlingual respeaking, aiming to shed some light and preliminarily give an answer to the following additional RQs:

RQ5: Is a master's degree in translation and interpreting an adequate environment to train students in intra and interlingual respeaking?

RQ6: Can respeaking training be useful in better educating students in Media Accessibility and, specifically, in DHOH accessibility needs?

All RQs are finally answered in the final part of this thesis "Conclusions". However, training details, workflow and experiment methodology relevant to RQ1, RQ2, RQ3, and RQ4 from EX1 are dealt with in Chapter 3, Section 3 and Chapter 4 > Section 3, while

relevant analyses are addressed in Chapter 5 > Section 3. Training details, workflow and experiment methodology relevant to RQ1, RQ2, RQ3, and RQ4 from EX2 are dealt with in Chapter 3 > Section 4 and Chapter 4 > Section 4, while relevant analyses are addressed in Chapter 5 > Section 4.

4. Thesis overview

This thesis is made up of five chapters in total, plus the introduction and the final part dedicated to the conclusions. In order to help readers, at the beginning of each chapter an introduction is provided to summarize the content of the following pages.

Chapter 1 – An introduction to respeaking introduces the basic concepts of the main focus of the research: the technique of respeaking. It introduces the more general field of AVT and explores the specific branch of MA, also providing some key information on subtitling. This first chapter is intended as the theoretical introduction to the thesis topic, covering LS and respeaking among other techniques, detailing history and previous literature on respeaking. The functioning of ASR and MT systems is briefly presented and then emphasis is posed on human-computer interaction as one of the focal points of the different methods analyzed throughout the work. Lastly, both intra and interlingual respeaking are covered, mentioning pioneering research in the field that made this thesis possible.

Chapter 2 – Training and assessment in interlingual respeaking aims at completing the theoretical framework needed to understand this work by dwelling on training in interlingual respeaking first, and on assessment in its last part. The chapter starts by presenting the ILSA project and other insightful research contributions on training models for interlingual respeaking. Then, assessment in live subtitles is tackled, covering linguistic accuracy and another important aspect: the delay in creation of subtitles. As far as accuracy

is concerned, different models to assess intralingual LS are presented (the WER and the NER models), to then introduce the NTR model for ILS assessment. An explanation of error grading and a digression on the different speech levels of analysis (dependent vs independent units) are presented since both were deemed crucial during the experimental phase. The chapter ends with some insights on assessment in SI, another critical point on which the whole research depends, highlighting the bias of subjectivity in translated, interpreted or respoken products.

Chapter 3 – Interlingual respeaking training workshops explores qualitative data that can answer to RQ5 and RQ6 about training in respeaking and educating for MA. The main experiment for the English to Italian workshop is described first, followed by the Spanish to Italian one. For both EX1 and EX2 participants, materials, teaching methods and training modules are detailed. In this chapter the first qualitative data gathered during the experiments is also presented, from the pre- and the post-workshop questionnaires. Particular emphasis is given to the participants' profiles and their previous knowledge and experience in MA, STTI and respeaking. By means of the questionnaires, feedback and thoughts on the overall training and practice experience are included in the chapter, reporting the students' considerations to help elucidate how much is known, or unknown, on LS and, above all, on respeaking among our interpreting students at the academic level.

Chapter 4 – Methodology and experiments settings is dedicated to exploration of the experiments. The methodology for both experiments is outlined in this chapter. After giving some information on similar research comparing different ILS methods, some suggestions on Machine Translation, as it is used for two methods in the second experiment, are proposed. For both experiments methodology is presented covering participants and teams, methodology and tested methods, software used, and tested chunks of text. Pre- and post-experiment questionnaires are also briefly discussed, aiming at detecting any technical issue encountered during the testing.

Chapter 5 – Results and discussion is where the final data extracted by the analysis are provided. Again, the structure first introduces the results from the main experiment and

then the results from the second, following a similar structure. In both cases, the NTR scores for linguistic accuracy are displayed first, followed by a detailed analysis of the different and most frequent error typologies detected in all outputs for translation and recognition errors for all the tested methods. Secondly, delay calculation averaged per method is presented and some final thoughts and discussion on the results obtained are outlined. After the same features for the second experiment are presented, the chapter concludes by comparing the results from the first three methods in EX1 and EX2, as well as comparing this research to the results obtained by the ones by the Universities of Vigo and Surrey. One last section is dedicated to training in respeaking for interpreting students.

The thesis is completed with a final section "Conclusions" in which answers to the RQs are given. Before answering the main RQ for the research, answers to RQ1, RQ2, RQ3, and RQ4 for EX1 and EX2 are presented. Afterwards, answers are also given to RQ5 and RQ6. In addition to bias of the study, possible further developments in this field are proposed, hoping to stimulate further research on the topic.

The research work also includes **9 appendices** where all materials collected during the two experiments are available. These include calculations concerning NTR analyses, pre- and post- workshop as well as experiment questionnaires and all video and audio recordings of the performances.
Experiment audiovisual materials and analyses are available on a dedicated cloud space at the given link, as well as on USB, if requested.
Workshop and experiment questionnaires are reported at the end of the thesis for readers, together with the reference cloud link that includes the participants' responses.

Chapter 1.

An introduction to respeaking

1. Introduction

This first chapter will offer a theoretical introduction with a top-down approach, i.e. from the general to the specific, to offer a comprehensive overview on the field of interest: ILS and, more specifically, respeaking. Having already provided some notions on accessibility in a broader sense, AVT will now be introduced with specific reference to MA, to then move to an introduction to pre-recorded and live subtitles, simultaneous interpreting and, ultimately, intra and interlingual respeaking. After laying the foundations of the two techniques that have been identified as the most skill-oriented for respeaking – namely subtitling and SI – an overview of useful expressions and terms to refer to LS is provided. Then, different methods to provide live subtitles are outlined with particular attention to respeaking and respeaking in Italy. Attention is then given to ASR and MT systems and how they work and a digression on Human-Machine Interaction (HMI) is provided, defining verbatim and sensatim subtitles. Concerning intralingual respeaking, a closer look at the different extents of HMI in the creation of LS is proposed through a small case study conducted at the beginning of this Ph.D. The chapter ends with a focus on interlingual respeaking specifying definition, skills and competences and process.

2. Audiovisual Translation

AVT is a rapidly growing and ever-evolving branch of translation which has also been tackled in Translation Studies (TS) in more recent years (Díaz Cintas, 2009; Pérez-González, 2008). Over the last fifteen years, demand for audiovisual translated products has increased considerably (Díaz-Cintas, 2009; Baños-Piñero & Díaz-Cintas, 2015; Pérez-González, 2019). As we move into a more technology-driven and media-oriented world, different AVT services are in great demand (Díaz-Cintas, 2009; Díaz Cintas & Remael, 2007, 2021) and in recent times have gained even more momentum with the advent of the Internet, social media and streaming platforms. Many definitions for AVT exist, and a great deal of research has and is currently being carried out in different areas of this field. AVT concerns "[…] the transfer of multimodal and mutlimedial texts to another language and/or culture" (Pérez-González, 2008). Therefore, with the expression AVT we refer to all modes of language transfer that aim to translate the original dialogues of audiovisual products, i.e. products that communicate simultaneously through the acoustic and visual channel, in order to make them accessible to a wider audience (Perego & Taylor, 2012). In Translation Studies, Jakobson's renowned taxonomy of translation (1959) seems all the more important in properly understanding the different approaches of AVT. Translation – in all its types – is indeed categorized as intralingual, where signs (words) belong and are transmitted from a language to the same language; interlingual, where verbal signs are replaced by other verbal signs in another language; and intersemiotic, where there is a shift from one code to another code, perceived through different channels (e.g., from the acoustic channel to the visual one, or vice versa). With the exception of dubbing, other AVT techniques imply an intersemiotic translation that can be either intralingual or interlingual depending on the source text (ST) and the target text (TT), which entails a transmission from different channels. Therefore, AVT by definition it aims at 'making contents accessible for a wider audience'.

2.1 Audiovisual Translation and Media Accessibility

Given AVT's very nature of 'accessible translation', it may be appropriate to dwell a little more on the specific aspect of accessibility in AVT (Díaz-Cintas, 2005).

As pointed out by Greco and Jankowksa (2020), Media Accessibility (MA) was initially mainly considered as a branch of AVT, and confined into Translation Studies (TS) field (Greco, 2019a). However, it can be argued that MA does not strictly belongs to the field of AVT, as with the increasing need for accessibility in almost all aspects of our lives a new interdisciplinary field has emerged over the years, the one of Accessibility Studies (AS), involving a vast number of areas (Greco, 2016; Greco, 2018; Greco, 2019b). Accessibility and the field of AS gradually entered the horizon of AVT, expanding over time to different modalities. AVT mainly aims to translate media contents from one source language to another target language through subtitling, dubbing and voice-over (Szarkowska & Wasylczyk, 2018), but it also expanded to audio and live subtitling, and sign language interpreting (Greco, 2018). Dawson (2020) provides an exhaustive overview of the emerging field of MA and its links with AVT, explaining how MA refers to AVT, either through the former including the latter (as per Szarkowska & Wasylczyk, 2018) or the two intersecting and working hand in hand (as per Greco, 2019a).

For the sake of clarity, although sharing common ground with AVT, AS appears to be the field to which MA properly belongs, since it has rapidly moved beyond the fields of AVT and TS. The branch of AVT does encompasses many different modes which imply a multimodal and multimedia transfer (Gambier & Gottlieb, 2001) to make media content accessible (Remael *et al.*, 2016), but according to Greco & Jankowska (2020) other modalities like audio description (AD) and Subtitling for the Deaf and Hard of hearing (SDH) pose a challenge in clearly distinguishing between the audiovisual and the MA fields (Remael *et al.*, 2016). AD for the blind and visually impaired (B/VIP) is a technique through which the visual contents not accessible to all are verbally described and explained aloud (Snyder, 2008; Fryer, 2017), namely where visual code is translated into acoustic; while SDH (and live subtitles) transcribe acoustic contents such as dialogues and sound features for those who cannot hear them shifting from the acoustic, oral input into written output.

Initially, MA exclusively concerned AD and SDH (Orero, 2004), mainly intralingual modalities of translation, to cater to the need of providing accessible audiovisual content for all, allowing subjects with blindness or visual impairment and hearing loss to access the same information as the rest of the audience. Nevertheless, "media accessibility has become a key concept in [AVT], devoted to studying how linguistic and sensory barriers can be overcome to make audiovisual products accessible" (Baños-Piñero, 2017: 485), which shows that over time there has been a shift towards interlingual interest as well, including subtitling, dubbing, and voice-over as well as SI. Given the stake taken that MA pertains to the independent growing field of AS, even more than to the AVT one, and therefore not only to Translation Studies (TS), the previous quote could be rephrased as "MA has become a key, fast growing field of interest in between TS and AVT, and AS, devoted to study how sensory and linguistic barriers can be overcome to provide multimedia accessibility for all" (*ibid.*).

3. Subtitling

Subtitling is one of the most prosperous areas in the discipline of TS and may be defined as:

> "a translation practice that consists of presenting a written text, generally on the lower part of the screen, that endeavors to recount the original dialogue of the speakers, as well as the discursive elements that appear in the image (letters, […] graffiti, inscriptions […] and the like), and the information that is contained on the soundtrack (songs, voices off)".
>
> (Díaz Cintas & Remael, 2021)

This definition sheds light on several interesting aspects of subtitling that will also be considered when specifically focusing on live subtitling. First, subtitling is a form of intersemiotic AVT, from oral input in L1 to written output in L2. Second, it is said that

subtitles are typically displayed on the lower part of the screen, but this depends on the position of the written text – 'sub' in "subtitling", for example, if displayed in the lower part of the screen, but 'sur' in "surtitling" or "supertitles" (Freddi & Luraghi, 2011) if displayed in the part above the stage, as in the Opera, or the screen. Concerning subtitles position on screen, in more recent years it has been argued that subtitling could – and should – shift becoming more to creative, thus allowing for original elements in subtitles not only in reference to the position on the screen, but to text font, colors, dimensions, and other aspects (Romero-Fresco, 2021). Lastly, subtitling in its broader sense can be intralingual or interlingual. Intralingual subtitling, also known as same-language subtitling (SLS), is a monolingual activity that is mainly associated with SDH. Interlingual subtitling encompasses discursive elements that appear on screen in L1 and are translated in L2, thus (sub)titled in the TL.

This last example casts light on the difference between subtitling and specific SDH: for an intralingual product of SDH, such discursive elements in the image would not be subtitled, as DHOH people would visually access that information. In a different case, namely the SDH version of audiovisual content in a different SL, that information would be included, broadening the accessibility from strictly sensory to linguistic. In other words, what has been described by Neves (2005) as a subtitling solution that implies the translation of messages from verbal and non-verbal acoustic codes into verbal and/or non-verbal visual codes, adapting them to the needs of people with hearing impairment to guarantee readability and thus greater accessibility.

This way, SDH can also be interlingual as subtitles for hearers, provided it contains specific sound features for non-hearing audiences, as better shown in Figure 2 below.

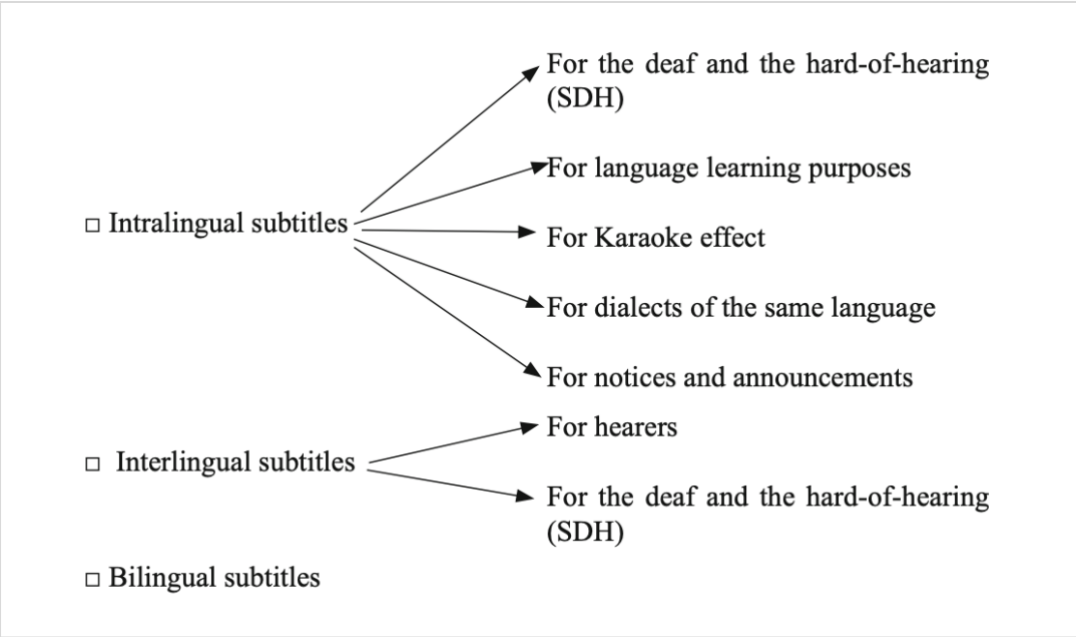


Figure 2 – Intralingual and interlingual subtitles overview (Díaz Cintas & Remael, 2007: 14)

Subtitles can also be classified depending on how (and when) they are created, whether that be pre-recorded, semi-live or live. Pre-recorded subtitles are created starting from the full script of a movie, video, or programs recorded in advance; they comply with the space and time constraints that are typical of subtitles, but they are created and adjusted to the changing of shot without the time pressures of a live version. Semi-live subtitles are created from the script of a live program which is available in advance, but usually shortly before the broadcast. This allows for the subtitles to be divided into chunks, but the timing is assigned during broadcasting since the subtitles are cued live, usually in blocks. Live subtitles are instead entirely created while broadcasting, for example, or in a live event which allows no time to adjust to shot changes due to simultaneous cuing and without any script available beforehand, something that is likely to increase delay in their broadcasting (Chmiel *et al.*, 2017).

3.1 Subtitling for the d/Deaf and hard of hearing (SDH)

As previously mentioned, SDH contains specific sound features to help DHOH audiences access the acoustic contents, and its demand is expected to increase (Romero-Fresco, 2019a). SDH – both intra and interlingual – is an intersemiotic translation (*ibid.*), from an acoustic code into a written one. This type of subtitling is carried out for programs that have been pre-recorded, thus respecting all the space and time constraints of subtitling but produced in a different moment. SDH is defined as:

> "a practice that consists in presenting on screen a written text that accounts for the dialogue and its paralinguistic dimension, as well as for music, sounds and noises contained in the soundtrack, so that audiences with hearing impairments can access audiovisual material".
>
> (Díaz Cintas & Remael, 2021)

Different guidelines are available for SDH, such as the use of labels or different colors to differentiate between speakers in frames where more than one person is speaking, or when someone not included in the framework (a voice off) is heard (RAI Italia, 2021). Music, noises, or other sounds in the soundtrack are usually inserted into rounded or squared brackets. Paralinguistic information is also underlined when necessary for appreciation of the message, including for example when utterances by characters are being stated with irony, particular emphasis, or specific accents.

3.2 Live subtitling

Live subtitling (LS) or speech to text interpreting (STTI) refers to the production of subtitles in real-time for live programs – sport matches, TV shows, the news, weather

forecasts, to name a few – or events such as conferences, meetings or parliamentary procedures. Unlike it happens for pre-recorded subtitling, due to the live production of captions LS is considered the most demanding form of subtitles. Furthermore, their real-time production allows for minimum review before they are broadcast (Eugeni, 2007). LS can be produced through different methods and with different approaches, among which respeaking is one of the newest. As pointed out by Díaz Cintas & Remael:

> "Traditionally professionals used stenotype or shorthand techniques or different keyboards to transcribe or translate the original dialogue, but these days respeaking, or speech-based live subtitling, is gaining ground in the industry".
>
> (Díaz Cintas & Remael, 2021: 10)

Different methods of LS are available varying from country to country, where some are more used than others. Allowing to produce a transcription of a live event (or an oral ST) with the shortest possible delay, the different techniques also depend on the setting and the context (Lambourne, 2006; Eugeni, 2008; Romero-Fresco, 2011). The methods traditionally used, for live transcription are:

- QWERTY
- Velotype
- Dual keyboard
- Stenotype
- Speech Recognition (respeaking)

Standard QWERTY keyboards are a simple way to transcribe using a keyboard and typing as fast as possible what is being said. This method is considered very slow (50 to 70 wpm

on average) (Romero-Fresco, 2011) and it is also highly exposed to typing errors due to the required rapidity.

Velotype is "a syllabic chord keyboard which allows the user to press several keys simultaneously, producing syllables and words rather than letters" (Romero-Fresco, 2011: 13), which allows for faster typing. Another method is the so-called dual keyboard, where two operators working in pairs and taking turns transcribe alternate utterances; a five-way QWERTY keyboard can also be used in some cases, allowing for up to five subtitlers to work on the transcription at the same time. According to the Encyclopedia Britannica online, stenotypy is "a system of machine shorthand in which letters or groups of letters phonetically represent syllables, words, phrases, and punctuation marks". Like velotype, stenotype is operated with both hands, and several keys may be struck simultaneously to print a complete word in one stroke". Stenotype allows for very fast typing, up to 250 wpm, and it is frequently used in court reporting, also in Italy. Lastly, subtitles can be produced through SR or respeaking. The technique of respeaking (*ibid.*) allows for production of subtitles by means of SR, as detailed by Eugeni (2008) and Romero-Fresco (2011). Despite the rapidity in creating the subtitles for stenotype, it is interesting to notice that according to the European Broadcasting Union (2004) "training a subtitler to respeak […] takes a couple of months; by comparison, training a stenographic subtitler from scratch can take 3-4 years", as pointed out by Romero-Fresco as well (2011: 15). Respeaking does seem the most time and cost-effective technique for live subtitles at the moment, and investments in training for this technique could be very fruitful. More needs to be detailed about its professional profile: it is argued that it can be easier for a subtitler or a simultaneous interpreter to perform in respeaking, as will be further discussed in Section 2.2 in Chapter 2.

As a point of reference, Romero-Fresco (2011) offers a very comprehensive overview of respeaking as a professional practice throughout many countries, detailing working conditions, methods, and best practices for respeakers in some European countries, as well as in Canada and in the United States of America. In the United Kingdom, respeaking for television was introduced in 2001 (*ibid.*) and it has quickly become the preferred method to live subtitle TV programs. Likewise, respeaking has quickly become more and more used

for live subtitles also in Spain, Switzerland, Flanders, Denmark, and France between 2006 and 2008, either introduced after dual-keyboard method, or directly used from the beginning of the service provided. In Canada, the first live subtitles though respeaking dates back to 2004, while for the USA stenotype has always been the preferred method for live or real-time captioning, the term used to refer to live subtitles in the US. The techniques of stenotyping and velotyping as well are much used in courtrooms throughout the country, to produce real-time transcription of trials and court sessions.

3.2.1 What to call it?

Previously it was pointed out how the term 'subtitling' itself would only be accurate for captions appearing in the lower part of the screen, while 'surtitling', for example, would fit for captions in the upper area, as for opera plays. As foreseeable, the terminology matter is also particularly varied for LS, lacking univocal terms to refer to the profession and, consequently, to the process. Drawing on what Eugeni and Bernabé (2019) outlined, there are several terms and expressions that refer to LS, as far as English language is concerned: 'live subtitling' or 'surtitling', indeed, sometimes proposed as '(sub)titling' (Romero-Fresco & Alonso-Bacigalupe, 2022) specifically for this ambiguity; 'live titling', which would allow to generalize the reference regardless the position of the subtitles on the screen; 'live captioning', or, again, the use of 'real-time' instead of 'live' for 'real-time captioning', 'real-time subtitling/titling'. 'Fast typing' is also a term which reference is made to when referring to the different methods outlined above, reflecting the rapidity needed in producing the transcript in live settings. Despite initially being referred to as intralingual service with DHOH subjects (Stinson, 2015), also 'speech to text interpreting' (STTI) is now also used interlingually to refer to the interpretation of an oral input into a written output via different tools: keyboard, stenotyping machine or SR software.

Although with full awareness of the importance of consciously using the different proposed terminology if needed, throughout this research it has been decided to use the term 'live subtitling' (LS) (or STTI) when referring in a broader sense to the field of interest, and 'interlingual live subtitling' (ILS), with specific reference to the same practice but carried out from one SL into another TL.

4. Respeaking

After having presented the different methods to produce LS in Section 3, the focus is here given to the technique of respeaking at live events (Moores, 2020).

Respeaking refers to the production of subtitles through SR systems. As a relatively new technique that is paving the way in this field (*ibid.*), it can be carried out both intralingually and interlingually. While intralingual respeaking is more widely known and employed to produce LS, interlingual respeaking seems at an earlier stage, from a professional and academic perspective. In addition to its English definition by Romero-Fresco (*ibid.*), a definition of respeaking in Italian based on the international investigation available at the date was first provided by Eugeni (2008):

> "Una riformulazione, una traduzione o una trascrizione di un testo […] prodotta dal respeaker ed elaborata dal computer in contemporanea con la produzione del testo di partenza […]. [Un] *software* di riconoscimento del parlato procede alla trasformazione dell'*input* orale in testo scritto[3]".
>
> (Eugeni, 2008: 7)

[3] "A reformulation, a translation or a transcription of a text […] produced by the respeaker and processed by the machine at the same time of the source text […] [in which a] speech recognition software turns the oral input into written text" (my translation).

In other words, to carry out a respeaking task, respeakers are involved in the repetition (or reformulation) of audio content, interacting with a machine to produce as a final output, a written text (subtitles), always bearing in mind that:

> "[…] the respeaker, far from being a mere parrot, has to perform a process of message comprehension and reformulation that often requires a certain distancing from word-for-word formulation".
>
> (Romero-Fresco, 2011)

4.1 What to call it?

As for the terminology related to the field of live subtitling (see Section 3.2.1 before), the term 'respeaking' is also subject to other contaminant names. Again, it may be useful to clarify from the beginning that 'live subtitling' is here being used as an umbrella term to generically describe the task, as per intralingual or interlingual STTI, while the term 'respeaking' is used to refer to the method to produce LS through SR (Dawson, 2020).
Drawing on what Romero-Fresco (2011) and Eugeni & Bernabé (2019) recapped, different terms and expressions can be retrieved in literature and previous research, although 'respeaking' – from the idea of 'speak again/repeat', 're-speak' indeed (Lambourne *et al.*, 2004) – seems to be the most widely used and consolidated (Romero-Fresco, 2008). More recently, though, it has been argued that this might not be the most relevant term for such technique, since it does not entail its interlingual declination (interlingual respeaking, indeed), from one language to another, better depicted instead by the term 'transpeaking' (from 'translation') (Pöchhacker & Remael, 2019). 'Real-time (or simultaneous) speech to text translation' (STTT) has been previously used to refer to an interlingual process in which from an audio input a written text is produced, either if created by a human or a machine (Russello & Prandi, 2021). As far as English language is concerned, even though English terms themselves have been widely used also in other languages, also 'speech-based live

subtitling' or 'SR-based subtitling' can be retrieved as alternatives, as well as 'real-time subtitling via SR' (Eugeni, 2008). As very interestingly pointed out by Eugeni & Bernabé, "this lack of univocal terms has led to restricted views of the profession and consequently of the training, which concentrates on some aspects only" (Eugeni & Bernabé, 2019: 87-88), such as the passage from language to another, the simultaneity of the process, or the position of subtitles appearing on screen but not on the technique as a whole. In their opinion, regardless the different methods and forms of written text activity involved in reproducing spoken discourse (respeaking, stenotype, velotype, etc.), they should all fall into the same category, being that for not limiting the profession to a technique, a context, or a target audience.

In Spanish, following the discussion presented in Romero-Fresco (2008), the term 'rehablado' has been recognized, while in Italian the term 'rispeakeraggio' has first been proposed by Eugeni (2006a) and has gained popularity together with 'respeaking'.

4.2 Origins of respeaking

As reported in Romero-Fresco (2019) the origins of respeaking date back to 1940s in a US courtroom, where it was proposed to repeat words of the source speech into a microphone, using a stenomask[4] to minimize noise (Mack, 2006a), and transcription was then made starting from the recording, since no SR systems were available at the time. It was in 1999 that the SR software Dragon Naturally Speaking was used for respeaking (Romero-Fresco, 2019b). The technique of LS through SR is currently being used in many countries to provide LS for a variety of TV programs, but to different extents: in some countries it is a widespread technique used in a variety of settings, in others it is still gaining momentum as it is one of the newest (Romero-Fresco, 2011). However, intralingual

[4] A stenomask is a hand-held microphone built into a padded, sound-proof enclosure that fits over the speaker's mouth, to allow a person to speak without being heard by other people, and to isolate the microphone from background noise.

respeaking is most known and employed, while interlingual respeaking – as seen, sometimes referred to as 'transpeaking' – can be considered still in its infancy (Dawson, 2020). With the beginning of the new millennium, intralingual respeaking started being widely practiced first in the UK (Lambourne, 2006; Romero-Fresco, 2011). In 2006 the First international seminar on real-time subtitling was organized and hosted by the University of Bologna (Eugeni, 2008), issuing on inTRAlinea[5] the proceedings of the seminar and starting a thriving research trend at an academic level in the field.

4.3 Respeaking in Italy

As pointed out in Romero-Fresco (2011), in the field of broadcasting Italy has been delivering live subtitles since 1988 with Colby, using Dragon Naturally Speaking to respeak, and touch screens to introduce punctuation, proper names and keywords. Research in respeaking in Italy was kick-started and pioneered by Carlo Eugeni since 2006, throughout the last fifteen years. Following his first research in the academic field, on respeaking for TV programs, respeaking technologies and perception, the technique has been capturing more and more attention also in fields other than the broadcasting one, like in events such as conferences, live meetings, parliamentary sessions and courtrooms among some of them (although velotyping is still very much used for transcription in this last context).

Nevertheless, respeaking cannot be considered a widespread technique quite already in our country. Specifically at an academic level, training in respeaking, both intra and interlingual, in Italy is lagging behind and offers very few opportunities. This is particularly of interest for this research given the interlingual respeaking workshops offered at UniGe to investigate the level of knowledge on MA and LS among master-level students, to try to answer the generic RQ1 and RQ2 of this research.

As many projects at a European level on STT-based tasks and respeaking are funded, the professional figure of respeakers has been gaining more and more attention in the industry.

[5] https://www.intralinea.org/index.php/print/article_specials/1684.

Even more, in the last two pandemic years in which almost all bulks of communications have shifted from in-presence to tele-communication online events, the need for subtitles produced live, and for transcriptions of dialogues while streaming has become a *sine qua non* condition for many. As a testimony, during this research more than one private agency got in contact to find out more about respeaking and seeking for professional respeakers for their online events.

The transfer to the online, always-connected mode of communication has interested all industries, from the business world via office videocalls and meetings, to the academic environment with online seminars, classes, and lectures. In the wake of this revolution, a quality-driven choice between human or machine-oriented methods for LS can become crucial for the industry. Interesting to notice, at last, that Italian as a SR language was still not offered in many of these platforms as per last year. It is also true that while intralingual respeaking has gained momentum, its interlingual declination is still in its infancy, in Italy and in other countries as well.

4.4 Subtitling and Simultaneous Interpreting in respeaking

As seen, respeaking as a form of live subtitling shares common ground with the technique of pre-recorded subtitling itself, and with SI. To some extents, respeaking can be compared to subtitling when the focus is on the product, and to SI when concentrating on the process (Russello, 2010). Over the last decade research in the field has bloomed, investigating more on the relationship between respeaking and both SI and subtitling, in intralingual but also interlingual respeaking (Romero-Fresco, 2015a; Chmiel *et al.*, 2017, 2017a; Szarkowska *et al.*, 2017, 2018, Dawson, 2020; Sandrelli, 2020), trying to research and find an answer to the question "Who are live subtitlers?" (Robert *et al.*, 2019a), and investigating their task through a competence-oriented lens (Pöchhacker & Remael, 2019). Through such recent research on the most suitable professional profiles to carry out respeaking tasks, subtitlers and interpreters appeared to be valid candidates (Szarkowska *et al.*, 2018; Dawson, 2020). In both studies, in an attempt to investigate the best suited profile

for respeaking, some participants with different academic and professional backgrounds were tested in respeaking tasks. Szarkowksa *et al.* point out that interpreters consistently achieved higher scores in output assessment than other groups, and Dawson highlights how, together with some other skills (such as knowledge on SDH and edition for subtitling, and short-term memory and multitasking for SI), a respeaker should also acquire specific skills such as dictation when interacting with the SR software. Overall, both interpreters and subtitlers were observed to be the best suited for the task, although some differences were noticed in performance ratings among the two groups. This allowed to shed light on the need for specific training for respeakers, drawing from both SI and pre-recorded subtitling. On the one hand, training competences of live tasks such as interpreting seems fundamental. In particular, for interlingual respeaking, it is interesting to notice the similarity between Gile's Effort Models (1992, 1995, 2015) in interpreting and the same short memory use in interlingual respeaking indeed. Also in intralingual respeaking, though, the cognitive load that respeakers undergo listening, respeaking while still listening, and while editing at the same time (Szarkowska *et al.*, 2017) is very important (Aliprandi & Verruso, 2006). On the other hand, training in subtitling covers some widely accepted guidelines such as characters/words per line, colors or tags to identify different speakers, duration of the display of the subtitles to be readable, among some of them. In addition, some task-specific skills for respeakers should be introduced and developed in training, as will be better detailed in following sections 2.2 and 3 in Chapter 2.

4.5 Automatic speech Recognition (ASR)

Respeaking would not be possible without the use of a Speech Recognition (SR) System. ASR allows human to interact with the machine using voice commands, either resulting in an automatic transcription of the audio input, or carrying out the command that has been given, the machine responding in a meaningful way thanks to a proper voice information processing, and recognition (just think of Siri on Macintosh products). There are several different ASR technologies currently available, but one of the most advanced is

the ASR based on Natural Language Processing (NLP), trying to resemble a natural conversation between the human and the machine, based on natural language creation and recognition. Let's take the case of using the ASR for respeaking, thus, to obtain the most possible accurate transcription of what is being said, as for respeaking interest.

All systems of ASR software, when listening to voice or audio inputs, break down words for analysis. As well – and very accessibly – explained by Pablo Romero-Fresco in the video lecture on SR and dictation of the ILSA course[6] (Module 2a, Unit 2), SR engines are normally made up of different models altogether, that are activated at different stages of the recognition process: an acoustic model, a vocabulary or a lexicon model, and a language model. An acoustic model consists in speech and audio data properly pronounced and spelt, together with their digital sound representation (i.e., how it is read and decoded by the engine). ASR these days not only extract the recognized text that was spoken, but they also interpret its semantic meanings through Deep Learning (DL).

At a basic level, when a sentence is dictated, the acoustic model digitizes it and breaks it down into phonemes, the basic building block sounds of words, but to do so, the soundtrack needs to be prepared and represented digitally. At this stage, the lexicon model is activated: it checks the recognized sound waveforms against the vocabulary, obtaining different possible words that can be recognized, due to assonance of same/similar pronunciation. To finally pick one word against another, the language model comes in. To do that, a language model consists of a corpus of texts that the speech engine can go through to analyze occurrences, collocations, and words position, statistically analyzing the probability of words being pronounced before or after the previous and next word, for example, and finally deducing the whole sentence as a meaningful unit. As pointed out by many professional respeakers, since the engine draws from the corpus of texts, the longer sentences are dictated, the better recognition is likely to be achieved, this way providing more context to the algorithm to cater for the language model.

[6] For more detailed information on how SR works, refer to: https://www.youtube.com/watch?v=JO5xcacxmAs.

The whole process depends on more modern, cutting-edge DL based on neural networks. In order to understand its basics, before DL, audio machine learning applications only depended on traditional digital signal processing techniques to extract sound features (phonetics and phonemes). With DL, the audio preparation for processing is set on standard data that allow to convert audio data into images, generating spectrograms of the audio (Doshi, 2021). Once the audio data are translated into spectrograms, this set of data is ready to be input in the DL model. Two main common approaches of DL architecture for ASR are used: CNN (Convolutional Neural Network) together with a RNN (Recurrent Neural Network) that demarcates each character of the words in the speech; or an RNN that reads each part of the spectrogram as one element in a sequence, thanks to an algorithm that takes character probabilities to detect their correct sequence.[7]

4.6 Machine Translation (MT)

"Speech recognition […] and machine translation are posited as a potential paradigm-changing language technologies with the capacity to disrupt the traditional configuration of the AVT industry" (Pérez-González, 2019: 11). Given the rise in the machine translation (MT) technology demand (Georgakopoulou, 2019: 526), two methods tested for the second experiment of this thesis involved the use of MT. Far from being one of the objectives of this research, which merely chose and integrated two ASR engines and one MT system in the methodology, MT will not be explained in detail in this section. The aim is only to briefly outline how it works, as proposed in the section above for ASR.

MT is the way to refer to automatic translation, in which a computer software carries out a translation of a text from a natural language to another. Basically, MT relies on artificial intelligence (AI) of the machine to translate contents without any human contribution. There are different types of MT: Rule-based Statistical Machine Translation (RBSMT),

[7] To look in more depth into audio DL for ASR, please refer to: https://towardsdatascience.com/audio-deep-learning-made-simple-automatic-speech-recognition-asr-how-it-works-716cfce4c706.

statistical MT (SMT) – also called Phrase-based Statistical MT, PBSMT –, and Neural MT (NMT).

RBSMT is the earliest form of MT, relying on the linguistic rules of the involved language pair according to many different bilingual dictionaries. This is something that requires human post-editing to great extent since it can often be of low general quality. SMT is based on the use of bilingual corpora containing billions of words in both languages, accounting this time to the relation between words, phrases and sentences, not just on the linguistic rules of a language. SMT draws from translated texts and align ST with TT lines on the basis of statistical calculations, also including data on occurrences in the TT (Georgakopoulou, 2013: 120). It was introduced by Google in 2006, and it is used by over 500 million users daily in more than 100 different languages (Romero-Fresco & Alonso-Bacigalupe, 2022; Georgakopoulou, 2019). NMT, being the most accurate MT system at present, has the advantage of using AI to train itself and therefore, it is always improving, as happens with a human brain when learning a new language. In 2016, Google itself shifted its SMT system to a NMT one (Google Neural Machine Translation, GNMT). Other providers followed suit, such as Microsoft just to provide one example, but for this research Google Translate was chosen as NMT software since it is the most widely used.

4.7 Human-computer interaction in ILS

This research is focused on ILS created through different methods that require different degrees of human-machine interaction. In 1992, Hewett *et al.* defined this interdependence with technology as 'Human-Computer Interaction' (HCI), or the implementation of interactive computing systems for human use. As explained in Eugeni (2019), in many fields today we see a rise in technological intervention rather than human which is gradually being replaced. In some fields where ASR and Natural Language Understanding (NLU) systems are used, for example, "technological evolution is reducing the place of humans in this interaction to such an extent, that their profession could hardly

be possible without it" in some fields (Eugeni, 2019: 873). Assuming that technological development is a crucial part of a translator, interpreter and respeaker's everyday life, it is necessary to further investigate if and how completely automated processes in this field can deliver acceptable standards, or under which specific conditions (Aliprandi *et al.*, 2014, Manetti, 2018).

When referring to different extents of HCI – or Human-Machine Interaction (HMI) – research has demonstrated that three different categories can be distinguished in ILS (Eugeni, 2019), that nevertheless can be easily applied both to intra and interlingual subtitling, given that this structure is technology-oriented (Pagano, 2020a).

The three categories are the following:

- Computer-Aided ILS, which implies a simultaneous translation by an interpreter and a live subtitler to transcribe what is being said; this form is mainly driven by professionals and the technology simply assists them in producing subtitles.

- Human-Aided ILS, which is the reverse of Computer-Aided ILS and in which the technology carries out the job and professionals edit the transcript created by the machine; here Automatic Translation (AT) software translates the input to different languages.

- Fully-Automated ILS, in which humans disappear. For intralingual LS, this is the case of automatic subtitles offered by web-based platforms (YouTube, for instance), where an ASR system transcribes the original and produces the transcript. For ILS, instead, this is the case of an ASR system recognizing the audio input in one language, connected to an AT software producing the transcription in a second language.

All three categories can be applied to respeaking as a method of LS, Romero-Fresco (2015) himself having defined respeaking as a form of computer-aided SI.

4.8 Verbatim and sensatim subtitles

As pointed out by Romero-Fresco (2009), the choice between delivering verbatim or sensatim subtitles has been controversial for years in the DHOH community. Verbatim subtitles are those delivered almost word-by-word and are supported and tend to be preferred by deaf associations and broadcasters as the best way to transmit the same ST content in AV products. Sensatim – or edited – subtitles on the other hand are paraphrased and therefore deemed, to some extent, as a form of censorship, since with the reformulation carried out by respeakers some information may be lost if compared with the information available to the hearing population (Ofcom, 2005: 17). Furthermore, verbatim subtitles are delivered faster than edited ones, since they require less effort by the subtitlers and, thus, are less expensive (Romero-Fresco, 2009). It must also be noticed that the debate between verbatim or non-verbatim subtitles also depends on the speech rate of the ST. As pointed out by Eugeni (2008), when a ST is particularly slow in terms of words per minute (wpm), the respeaker could and should aim for a verbatim rendering of the source in order not to slow the subtitles broadcasting too greatly; on the contrary, when the speech rate scales up, some reductions and condensations will be needed. Such manipulation of the ST affects orality features that are very typical in live speeches, such as repetitions and redundancies, self-corrections, hesitations, false starts, or parenthetical elements (Eugeni, 2008; Pagano, 2020a). When assessing the quality of verbatim subtitles the issue of speed (speed of text on screen) is crucial in the evaluation. Depending on the program broadcasted, speech rates differ and so does the speed of the subtitles: verbatim subtitles can be delivered faster since no reformulation is required, but does 'fast' automatically mean 'better quality'? Edited subtitles are slower and can indeed be considered by many as 'better' since they are read more easily (Romero-Fresco, 2009), although of course this is subject to the reading rate of the audience.

4.8.1 Human-machine interaction in intralingual respeaking[8]: a comparative case study

A first attempt of analyzing outputs created through different methods of live subtitling with different degrees of HMI covered intralingual respeaking. The focus was on live subtitles created intralingually in a Court Reporting setting, specifically for a Rome City Council meeting on November 12th, 2019. Through the Tiro project, a project by the Municipality of Rome (Eugeni, 2019), the Council sessions were made accessible in real time to d/Deaf people though intralingual live subtitling (Italian to Italian) and Italian Sign Language (LIS) Interpreting. The live subtitles were produced via two methods: the first method was respeaking and live editing, in which an intralingual respeaker produced the subtitles and a live editor corrected possible mistakes on the fly; the second method was through ASR and live editing, where the transcript was carried out automatically by the software taking the input from the speaker's microphone, and an editor monitored the subtitle before it was sent on-air. In doing so, sensatim subtitles, namely paraphrased subtitles, were delivered through the first mode of respeaking, while nearly verbatim subtitles were produced through the second mode: in respeaking, indeed, it is rare that verbatim subtitles are produced since editing leads to a minimal loss of information (Romero-Fresco, 2009). Respeakers were two and worked in pairs in a sound-proof booth, taking turns one carrying out the respeaking task and the other the editing, and vice versa.

In this first approach to the subject the linguistic analysis was not carried out using a model (refer to Sections 4.2 and 4.3 in the next Chapter 2 for linguistic analysis models in intra and interlingual respeaking). The Parliamentary session was examined linguistically focusing on the features of the different speakers and strictly offering a qualitative perspective on the results obtained. What is interesting to notice is that even from this preliminary approach that undoubtedly needs more comprehensive research, some differences did emerge between the two methods of delivery. As pointed out in Pagano (2020a) and Eugeni

8 Some parts of this section have been published in Pagano, A. (2020a) 'Verbatim vs. Edited Live Parliamentary Subtitling' in Dejica, D., Eugeni, C., Dejica-Cartis, A., *Translation Studies and Information Technology – New Pathways for researchers, teachers and professionals*, Editura Politehnica Timisoara, pp. 32-44.

(2021), generally speaking live subtitles produced through ASR and live editing are verbatim, unless the live editor decides to omit something, which is rarely the case because of the short amount of time available. Syntactically, sentences remain as complex as the original, and hence are sometimes more difficult to read since they also tend to be faster. On the contrary, live subtitles through respeaking are normally reformulated, unless the input is syntactically good. The ability of the respeaker to reformulate when necessary achieves an overall more coherent and readable text with shorter streamlined sentences and without ineffective orality features. Concerning delay, reformulation in respeaking leads to an average delay of just a couple of seconds more when compared to automatic subtitles, but it allows for more coherent punctuation.

In conclusion, at first glance the respoken intervention in this case does improve the overall quality and readability of subtitles in live settings where fast speech rates, several features of orality, and complex syntax are predominant, and reformulation is fundamental in reducing both the delay caused by the machine workload and the live editing process. Furthermore, reformulation aims at reducing the time and energy needed to read and understand the load of information available on screen.

4.9 Intralingual respeaking

Although this research focuses on interlingual respeaking and other live subtitling methods – still from an L1 to an L2 – it is relevant to provide some context for intralingual respeaking tasks as well, since the vast majority of skills and competences are shared among the two types of respeaking (ILS processes and competences will be presented in Section 4.10).

Different topics have been researched concerning intralingual respeaking, such as the differences between stenotype and other methods of real-time subtitling compared to respeaking (Eugeni & Mack, 2006), skills involved (Remael & van der Veert, 2006), respeaking vs. shadowing and SI (Eugeni, 2008), quality calculation (Eugeni, 2012)

intralingual respeaker training (Arumí Ribas & Romero-Fresco, 2008; Romero-Fresco & Martínez, 2015), respeaking compared to other techniques for specific genres (Matamala *et al.*, 2017), and respeaker professional profile, to mention a few. In an attempt to define a competence-oriented profile for respeakers, specific attention is given to the process required by respeaking, and the different skills involved. The process in respeaking, as already defined (*ibid.*), is not carried out all at once, but is a multi-step process developed through different stages, in which the respeaker:

> "listens to the audio input through a headset and uses a microphone to rephrase this input to a computer with respeaking and subtitling software, which together turn the spoken input into written subtitles […] [that are] edited using the computer keyboard before they are broadcast".
>
> (Pöchhacker & Remael, 2019: 134)

In light of the labelling identified by Gile in his interpreting Effort Model (Pöchhacker, 2008), Pöchhacker & Remael (2019) identify different competences required in respeaking. Despite their proposal focuses on ILS (transpeaking), namely from one natural language to another language, some are shared by intralingual respeaking as well such as listening comprehension, strategic reformulation, dictation, and monitoring, thus it is a multitasking process involving different skills, sometimes at the same time, requiring great coordination and control.

4.10 Interlingual respeaking

Interlingual respeaking is gaining momentum as more inclusive live events take the lead. As in the intralingual mode, interlingual respeaking is a multi-step process that involves different skills and competences, even more so when translating between L1 and

L2. Therefore, interlingual respeaking is considered a quite complex and highly demanding task, to such an extent that research was needed to test its feasibility (Dawson, 2019; Davitti & Sandrelli, 2020). After concluding that interlingual respeaking is indeed feasible and having investigated the best suited profile for the task (Dawson & Romero-Fresco, 2021), in an attempt to thoroughly inform respeakers' training design, the focus was placed on which competences the professionals had to develop and train. As Pöchhacker & Remael point out in outlining the process of transpeaking:

> "[W]hereas the transpeaker will probably need to cope with a high audio input rate, the speed of TT presentation is constrained by the processing capacity of the software and by the target audience's reading-time needs, and the TT's physical form depends on the space available […] [resulting] in the need for strategic compression".
>
> (Pöchhacker & Remael, 2019: 135)

Therefore, among other factors of difficulty contributing to a higher cognitive load for the respeaker, speed is very important. It could be argued that is even more crucial in LS than in SI, and even more in respeaking than in steno or typing techniques since it entails extra time in the creation of the subtitles which is dedicated to the information processing by the voice recognition software. In addition, a respeaker needs to dictate punctuation and speak very clearly (dictate) for proper SR. Lastly, in interlingual respeaking the major challenge is to activate different skills throughout different steps, and some of those at the same time, since respeakers focus on ST comprehension and translation while continuing to listen to the audio input, monitoring and editing the TT if necessary before the subtitle is broadcasted.

We will now take a closer look at the ILS process as a whole.

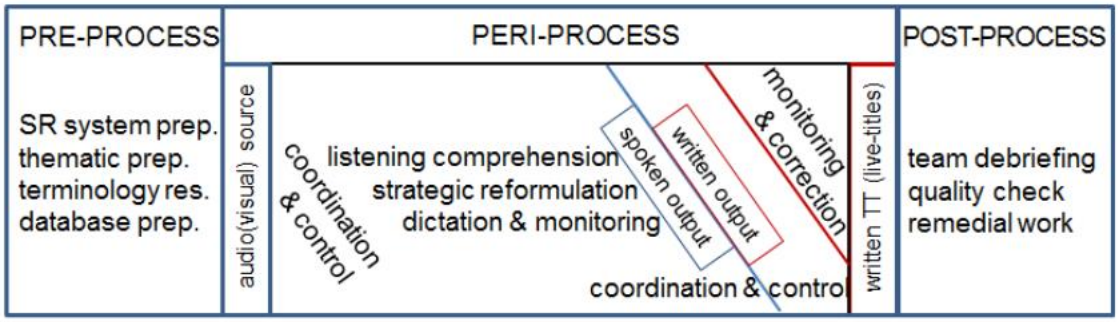


Figure 3 – ILS process model (Pöchhacker & Remael, 2019: 135)

As can be observed in Figure 3, three stages have been identified: pre-process, peri-process and post-process. The pre-process, prior to carrying out the respeaking task, is the stage that refers to the methodology research and SR software preparation. The peri-process coincides with effective development of the respeaking task (from spoken input to written output), activating listening comprehension in ST, strategic reformulation and translation in TT, rendering it through dictation, while monitoring the output. The post-process is the stage required after completion of the task in search for possible improvements, and it is the moment where quality of the output can be assessed.

Recent research conducted by Hayley Dawson (2020), identified five task specific skills for interlingual respeaking thanks to an experiment where different participants and trainees were involved: multitasking, dictation, live translation, language, and comprehension, reduced to the first three as proposed by Dawson and Romero-Fresco (2021). Similar skills were identified also by Russello (2010) as multitasking skills, live skills and delivery skills.
Pöchhacker & Remael (2019) first identify five general competences in the process: linguistic and cultural competence, world knowledge and subject-matter competence, technical-methodological competence, (inter)personal competence, and professional competence, that are required throughout the whole process of interlingual respeaking. In addition, six sub-competences specifically belong to the peri-process stage: preparation on the interlingual respeaking task and process, research and preparation, translation, multitasking, audiovisual monitoring, and editing.

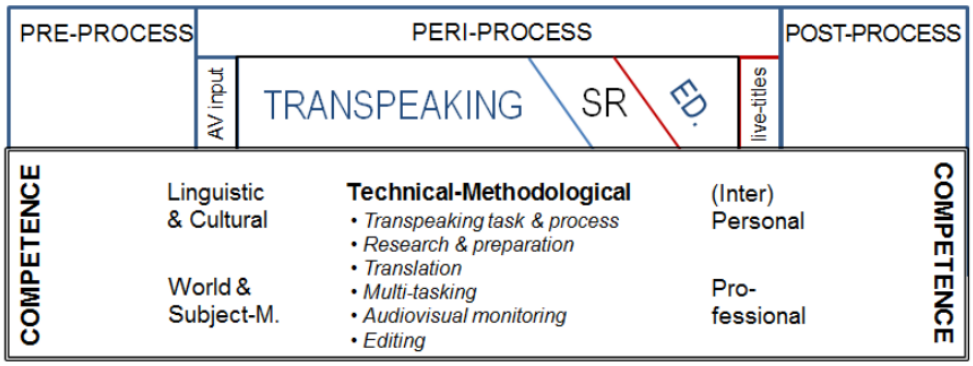


Figure 4 – ILS process and competence model (Pöchhacker & Remael, 2019: 137)

It is interesting to notice how, according to Dawson, apart from the five task-specific skills above mentioned, other skills deriving from subtitling and interpreting are deemed important in the interlingual respeaking process, such as working at speed, split attention, TL expression, segmentation, error correction (editing), short-term and long-term memory, reflection, and critical analysis through the different stages (Dawson, 2020: 206-219).

4.10.1 Pioneering projects

At this stage, after having introduced the fundamentals of ILS and respeaking and having described some of the most cutting-edge research in the field during the last decades, it is relevant to mention some recent projects that have paved the way for future professionals and future research in ILS. This research project itself finds its foundations in some of the findings and data that derive from these projects, thus it is all the more important to mention them. They both form part of the GALMA research group (Galician Observatory for Media Accessibility)[9], launched by the University of Vigo (Galicia, Spain), and aimed at promoting accessibility and translation in AV media through research, training and practice.

[9] http://galmaobservatory.eu/projects/galician-observatory-for-media-accessibility-galma/.

While this research was first conceived and started, one of the projects co-funded by GALMA, the ILSA (Interlingual Live Subtitling for Access) project[10], was about to finish, with less than one year of three left (2017-2020). More on this Erasmus+ EU co-funded project will be detailed in Chapter 2, since its intellectual outputs on interlingual respeaking feasibility, professional profile, and training model, among others, largely contributed to the hereby presented experiments. The project also tested and validated the first online course on interlingual respeaking.

The SMART (Shaping Multilingual Access Through Respeaking Technology) Project[11], together with SMART2, (2017-2019) aimed at investigating whether a specific training background can support the acquisition of skills needed to perform interlingual respeaking, feeding existing research on similarities between this challenging task and SI.

Of further interest is the LTA (Live Text Access), another Erasmus+ EU co-funded project[12] (2018-2021) aimed at offering proper training for live subtitlers in an attempt to harmonize the professional profile across Europe and create open-source material on LS. The project also created an online course of LS where both respeaking and velotyping are presented.

10 http://galmaobservatory.eu/projects/interlingual-live-subtitling-for-access-ilsa/.

11 http://galmaobservatory.eu/projects/shaping-multilingual-access-through-respeaking-technology-smart/.

12 https://ltaproject.eu/.

Chapter 2.

Training and assessment in interlingual respeaking

1. Introduction

In this chapter, further theoretical insights are given on the specific field of ILS focusing on two main aspects: training and assessment. Chapter 2 opens with an overview of the Erasmus+ project ILSA by the University of Vigo, in which the researcher of this thesis shortly participated in its final stage for the creation of some materials for the interlingual respeaking online course. The ILSA project was aimed at bridging the gap in interlingual respeaking training, proposing a course based on a training model. Comparison between the ILSA course proposal and the research-informed training model by Dawson (2020) in her Ph.D. research are presented, since the training workshop at UniGe, presented in the following Chapter 3, drew upon both. After training, assessment in LS is tackled. After reviewing the complex concept of 'quality' and defining delay as one crucial factor to consider in assessment, models are presented for the evaluation of intra and interlingual LS. For intralingual LS, this comprises the WER (Dumouchel *et al.*, 2011) and the NER (Romero-Fresco & Martínez, 2015) models. Finally, ILS assessment is illustrated through the NTR model (Romero-Fresco & Pöchhacker, 2017), with particular attention to the issue of subjectivity in assessing translated and interpreted texts as being one of the most complex tasks.

2. The 'Interlingual Live Subtitling for Access' project

At the end of the previous chapter some innovative and cutting-edge projects were briefly listed, which are yielding tangible results on dissemination in the field across the European Union. The ILSA project was the main source of information for this research, thanks to the creation of a training model for interlingual respeaking.

The acronym ILSA stands for 'Interlingual Live Subtitling for Access'. ILSA is a three-year Erasmus+ project co-financed by the European Union from 2017 until 2020. The project coordinator was the University of Vigo, Spain, and the other European partners involved were the University of Antwerp, the University of Warsaw, and the University of Vienna as well as other bodies. The key objective of the project is to bridge the gap between intra and interlingual live subtitling as recognized professional practices (Robert *et al.*, 2019b) by identifying the profile of the interlingual live subtitler and developing the first training course on ILS. In addition, the project outlines a protocol for the implementation of ILS in three scenarios: TV, live event settings, and the classroom environment. As an EU project, its impact was monitored and had a successful effect in several countries around the world regarding training, research, professional practice, and legislation. Among different deliverables and handouts throughout the journey, the ILSA project gave life to several intellectual outputs and much research in the field of interlingual live subtitling. The project prompted a large survey offering an overview on the state of intra and interlingual live subtitling around the world and provides guidelines on how to implement intra and interlingual LS in different settings and completed experiments that are crucial for the purpose of this thesis. Some of the first ever experiments comparing the performance of subtitlers and simultaneous interpreters in ILS were carried out in detail (Dawson, 2019; Robert *et al.*, 2019a) to help in defining the skills and competences needed for the professional interlingual respeaker. Through a competence-oriented task analysis of ILS, a proposal for the interlingual live subtitler profile was defined (Pöchhacker & Remael, 2019). This doctoral research rests largely on these foundations, without which it would not have been possible to carry out other experiments in the field. Starting from the professional profile and the training model identified through research-informed investigations, the

respeaking workshops at the University of Genoa itself were designed, as will be detailed in Chapter 3.

2.1 The ILSA course

Among other materials, presentations, and scientific publications, a milestone for the project was the design and publication of the first online and open-access course on ILS. The course is free and available on the University of Vigo's Campus do Mar MOOC platform, and is structured by seven different modules.

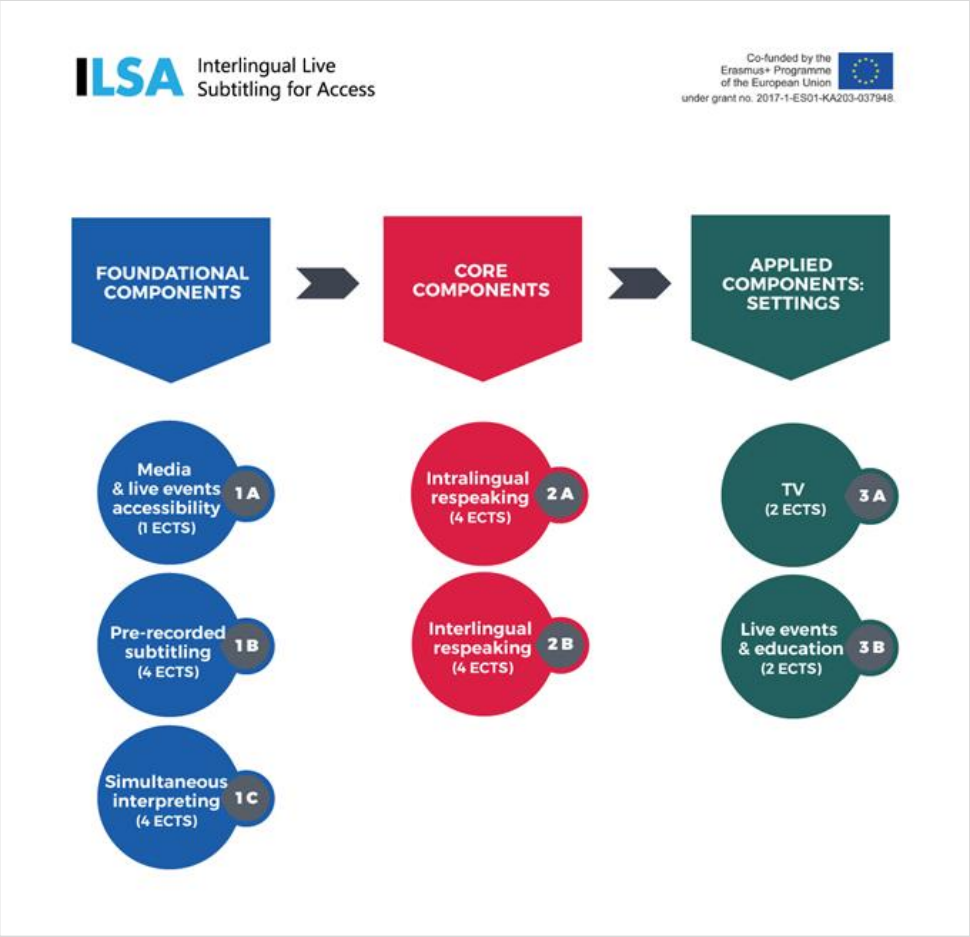


Figure 5 – The ILSA course structure

Based on the findings of the professional profile and the validation of the training model for interlingual respeaking, as can be observed in Figure 5 the course amounts to 21 ECTS (European Credit Transfer System), and develops through three main parts: foundational, core and applied components. The whole course is delivered self-paced and contains video lessons, reading tasks, quizzes, and guided exercises for practice. The course is offered in English (articles and references, tasks, and exercises, etc.) but the practice can be carried out in any language directionality by students, since technical and professional advice is given regardless of the TL. The foundational components are structured via three modules: Module 1a) "Media and live events accessibility", Module 1b) "Pre-recorded subtitling", and Module 1c) "SI": after a first theoretical introduction to MA, a review of the two tasks of subtitling and SI is offered to prepare students for respeaking. The core components gradually aim at introducing and practicing respeaking: Module 2a) "Intralingual respeaking" and Module 2b) "Interlingual respeaking" tackle the main part of the course, detailing pre, peri and post process of both tasks and training intra and then interlingual respeaking for their application in specific settings. The applied components are modulated through Module 3a) "TV" and Module 3b) "Live events and education", where beginner, intermediate and advanced levels of exercises are suggested.

2.2. A research-informed training model for interlingual respeaking

As pointed out by many researchers so far, over the past two decades intralingual LS has become a widely used professional practice and is recognized at an academic level (Eugeni & Bernabé, 2019; Robert *et al.*, 2019b; Dawson, 2019; Pöchhacker & Remael, 2019). Intralingual LS through respeaking, more specifically, has been practiced from 2001 (Romero-Fresco, 2011) and the interest in the professional profile of intralingual respeakers grew not long afterwards. Training for professional respeakers is not widespread and lacks a straightforward profile for the task, both intra and interlingually. Respeakers' training

models were previously proposed (Arumí Ribas & Romero-Fresco, 2008; Romero-Fresco, 2012), helping to inform the skills and competences they needed for the job.

Nowadays at a European level very few universities offer courses on intra or interlingual LS, and if so, they are only modules in AVT or conference interpreting master's degrees, thus mirroring a need for better training. Among them, the University of Antwerp, the University of Warsaw, and the University of Roehampton offer formal training in respeaking from 3 to 6 months, while the Universidade de Vigo offers a five-month online module in Intralingual Respeaking in English, Spanish and Galician and a three-month online module on Interlingual Respeaking in the same languages (Romero-Fresco, 2018). The University of Leeds, the Universitat Autònoma de Barcelona, and the University of Parma (Italy) provide introductory workshops on respeaking, and from 2020 the University of Genoa also offered a very first introductory session by the researcher of this thesis on intra and interlingual respeaking for the sake of this thesis (refer to Chapter 3 for more on this). In the EU are also worth mentioning the School of Applied Linguistics of the Zurich University, and the introductory course on respeaking at the University of Mons (Eugeni & Bernabé, 2019).
Different projects are also seeking to bridge the gap between training and the profession in real-time intralingual subtitling in general (LTA project; Eugeni & Bernabé, 2019) and between the professional practice of intralingual and interlingual respeaking in ILS (ILSA project).

A research-informed training model for interlingual respeaking was first proposed by Dawson (2020) through Doctoral research that intertwined with the ILSA project, although they differed in purpose, structures, and modules. As described above, the ILSA course on interlingual respeaking was conceived to be integrated and provide credit in Higher Education (HE) courses, providing open-access learning resources and materials that can be selected and used by trainers and trainees. The training model "can be used as a guide to develop future interlingual respeaking training" (Dawson, 2020: 231). In other words, the research-informed model can help design a training to be integrated as part of a master's program, or adapted to other needs of modules on respeaking. Furthermore, the ILSA

course and the research-informed training course (Dawson, 2020) differ in module structures, although foundational components of MA, subtitling, and SI have been included in both of them. Another interesting difference is the one pointed out by Dawson herself concerning the introduction of the SR software, which is introduced as a 'Dictation and software management module' in the research-informed training course from the very first week of training, while in the ILSA course it is only introduced in the intralingual respeaking module (core components). It seems relevant to underline this, since one of the major issues for participants in this experimental workshop and research when respeaking was precisely dictation and segmentation and the management of the SR software.

The research-informed model for interlingual respeaking contributed to a research-informed training course that consisted of five modules, each containing different Units, and was conceived to last one semester (24 weeks). Figure 6 details the structure of the proposal: Module 1) "Media Access" (including subtitling); Module 2) "Dictation and software management"; Module 3) "Simultaneous interpreting"; Module 4) "Intralingual respeaking"; and Module 5) "Interlingual respeaking". The course also entails two discussion points on the professional world of LS and new developments in respeaking.

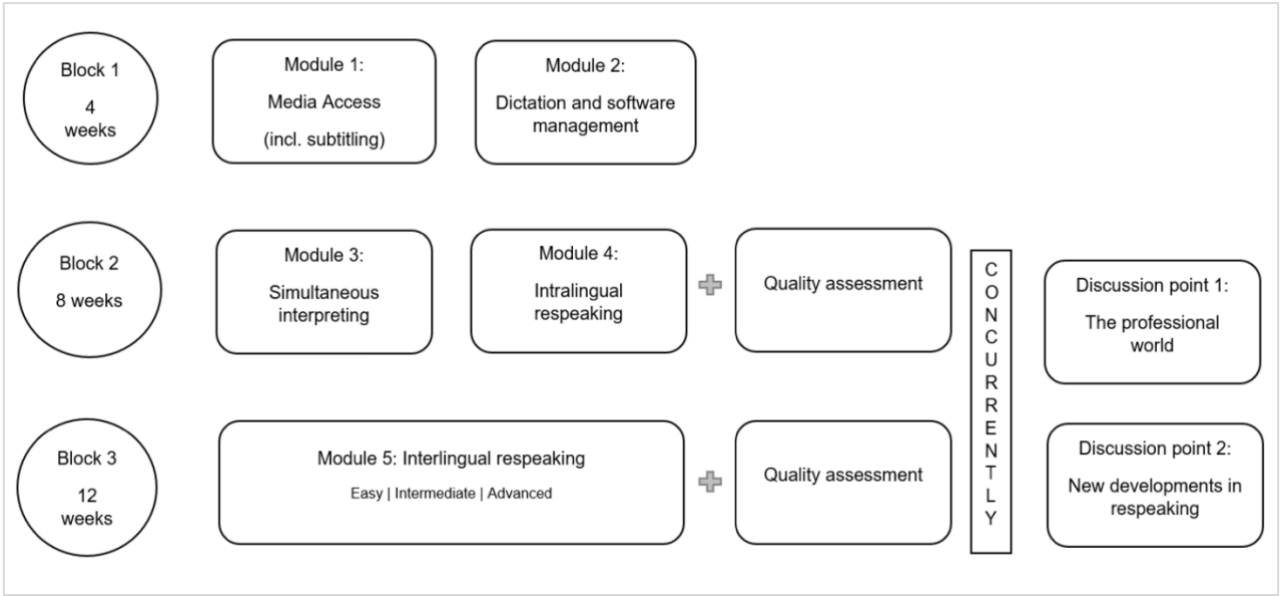


Figure 6 – The research-informed training model (and course structure) for interlingual respeaking (Dawson, 2020: 224)

As can be observed by comparing the two courses, the structure of the research-informed course by Dawson and the ILSA course do signal a gradual ascent towards interlingual respeaking, both offering practical exercises and activities on the major topics of MA, subtitling, SI, and intralingual, respeaking, while differing from module topics and exercise typologies in some cases (the research-informed training course specifically placing more emphasis on the progression of basic, intermediate and advanced practice).

For several reasons, the very first training proposal at the University of Genoa on interlingual respeaking drew from both of the discussed training courses. We will now go through an overview of the topics and modules tackled in the workshop developed for this experiment.

3. Training modules in interlingual respeaking workshop

For the purposes of this research, the first ever introductory respeaking workshop was offered at the University of Genoa (Italy) from English to Italian during the academic year 2020/21, in the Department of Modern Languages and Cultures where the researcher carried out their PhD. The workshop training materials for each module, as well as its questionnaires and activities will be described in detail in Chapter 3, Section 3. A second edition of the workshop from Spanish to Italian was then offered the following year (academic year 2021/22) as detailed in Chapter 3, Section 4. The different modules that were conceived drawing from both the research-informed model and course and the ILSA project course will be listed here to provide an overview of the structure of the workshop.

Firstly, it was named a 'workshop' since it was not a formal training course due to time restraints: the only viable proposal was to offer an introduction on interlingual respeaking, thus without the chance of emulating either a 6-month (researched-informed), or 21 ECTS (ILSA) course. The training lasted three months and was mainly inspired by the structure of the ILSA course, however it drew upon some aspects of the structure of the research-informed model and course too. In accordance with the competence-oriented model

(Pöchhacker & Remael, 2019) the participants were students and practitioners from the master's degree course in Translation and Interpreting, thus already pertaining to the world of AVT, especially for what concerns SI while few of them had previous experience in subtitling.

The first workshop was offered in academic year 2020/21, from English to Italian and was made up of four modules including the introduction on "Media and live events accessibility": Module 1) "Introduction to pre-recorded subtitling"; Module 2) "Simultaneous Interpreting"; Module 3) "Intralingual respeaking and SR software"; Module 4) "Interlingual respeaking". After the completion of the workshop and the completion of the experiment, analysis of the different respeaking outputs highlighted some gaps in trainee knowledge where room for improvement in training was available and most necessary. Since participants were students with very similar backgrounds and similar levels and competences in both SI and subtitling, when a second edition of the workshop was proposed one year later, its design was slightly modified. As noted, dictation to the SR software and sentence segmentation were identified as major issues when respeaking as part of the experiment. Therefore, as suggested by Dawson, in the second workshop training focused more on this area and SR was introduced sooner than in previous training experience, "as it is hoped that early dictation practice […] will ensure trainees' voice profiles are trained to a good standard before they begin respeaking practice" (Dawson, 2020: 232), as well as giving the chance for trainees to enhance dictation and segmentation skills. The second interlingual respeaking course/workshop for the academic year 2021/22, this time offered from Spanish to Italian, combined the introduction on MA into a module with subtitling and was structured with five modules in total: Module 1): "Media accessibility + pre-recorded subtitling"; Module 2) "Simultaneous interpreting"; Module 3) "Automatic SR software"; Module 4) "Intralingual respeaking"; Module 5) "Interlingual respeaking".

4. Assessment in live subtitling

After having introduced the basic concepts of respeaking and having analyzed the different approaches to train professional respeakers using research-informed methodology, we will now move to the concept of assessment in live subtitling, first focusing on intralingual LS, and then on ILS.

Assessment is a fundamental part of respeaking, as it is particularly useful to evaluate the final product of a respeaking task when aiming for best practice and searching for improvements. Thinking back to the different stages of the respeaking process (pre-, peri- and post-process, *ibid.*), the post-process describes when the assessment is carried out. Assessing the quality of the subtitles produced allows respeakers to identify mistakes, improve SR in the software by adding and training new words, and by highlighting strategies and good solutions that were implemented (Pöchhacker & Remael, 2019). It has been also pointed out that carrying out a quality assessment can be very useful during training since it gives trainees the chance to better spot any gaps in their knowledge and make plans of action for improvement where most needed (Dawson, 2020). Another important factor that determines the quality of LS is speed with which subtitles are displayed and how long they remain on screen to be read (Romero-Fresco, 2011; Davitti & Sandrelli, 2020), although this will not be calculated as part of this research.

4.1 Delay and the concept of 'quality'

'Quality' is not always a simple and straightforward concept to identify. The 'high quality' or 'low quality' judgment on almost everything is heavily driven from subjective elements such as personal opinions and preferences, for example, and it is therefore difficult to standardize with clear-cut characteristics.

Similarly to translated and interpreted products, quality assessment in LS cannot be restricted to mere linguistic accuracy, but includes other factors (Pöchhacker, 2013: 34)

such as delay. For example, in SI, *décalage* plays an important role in assessing the overall quality of a SI output, and reformulations are often the case when carrying out a simultaneous interpretation. More on subjectivity in interpreting will be tackled in Section 4.5, as it was particularly relevant to the analyses carried out for that research.

Delay is deemed even more important in LS than in SI – perhaps with the exception of media interpreting –, and possibly more important in respeaking than other LS methods since there is extra delay once the text has been respoken, i.e., the delay of the SR engine in processing words (*ibid.*). Between accuracy and delay there is an important link, underlined by Romero-Fresco, who points out that:

> "[T]he interplay between accuracy and delay constitutes an intrinsic part of live subtitling and is often described as a trade-off: launching the subtitles without prior correction results in smaller delays but less accuracy, while correcting the subtitles before cueing them on air increases accuracy but also delay".
>
> (Romero-Fresco, 2019b: 99)

Delay in subtitling is often referred to as 'latency', i.e. "the delay between the source speech and real-time target text, [which] will also vary in relation to the output delivery and degree of editing" (Davitti & Sandrelli 2020: 105).
As described by McIntyre *et al.*, 2018 (in Moores, 2020):

> "there is an inherent delay or latency in respeaking between a word being spoken and appearing on screen in a subtitle; this results from the time needed for the spoken word to be heard, respoken, recognised and processed through the subtitling software and onto the screen".

It is indeed important to calculate how many seconds the subtitles take to be displayed from when the ST is spoken (again, the shorter the delay, the better, as long as good accuracy is also provided). To conclude, an overall quality assessment that takes into account not only linguistic accuracy, but also latency among other factors, is the one that was sought while analyzing the outputs of this experiment and guaranteed the most comprehensive assessment.

4.2 Assessing intralingual live subtitling

Different models have been designed and tested over the years to assess quality in intralingual LS, regardless of the method by which they are produced, and especially for television programmes. In the following section we will focus on two word-based models for accuracy in intralingual LS, the WER (Word-Error-Rate) model, the basics of which laid the foundations for the NER model (Romero-Fresco & Martínez, 2015), the latter introducing error weighted categories (minor, standard, major) to insertions, substitutions and deletions, and for both recognition and edition errors. As regard to conceptual measurements Another model developed to evaluate quality of intralingual LS is the IRA (Idea Rendition Assessment) by Carlo Eugeni (2017), based on the main distinction of rendered and non-rendered ideas in the TT, and therefore more on communicative and conceptual level of analysis based on ideas, more than taking into account formal errors. Based on these two evaluation models mainly used for LS on TV (NER and IRA models), Eichmeyer-Hell (forthcoming) develops in her Doctoral research the WIRA (Weighted-Idea-Rendition-Assessment) model, including an extra category of 'very minor errors', and of 'partially rendered'.

The assessment of subtitles in respeaking, as pointed out, allows respeakers to identify mistakes and identify room for improvement, spotting their main areas of weakness (e.g. for dictation, carrying out an accurate final analysis of output can help highlight both how and which words should be dictated for better recognition). In respeaking, assessing both

human intervention (edition errors) and technological performance (recognition error by the SR engine) is necessary (Dawson, 2020).

4.2.1 The WER model

As pointed out in Romero-Fresco and Pöchhacker (2017), many assessment models for intralingual LS are based on the WER (Word-Error-Rate) model, used to analyze accuracy in SR-based products (Dumouchel *et al.*, 2011). To calculate accuracy, the model accounts for three types of error: deletions, substitutions, and insertions. The sum of these three error categories is subtracted from the total number of words (N), divided by the same number of words, and then multiplied by 100 to obtain the reference percentage.

$$\text{Accuracy rate: } \frac{N - \text{Errors}}{N} \times 100 = \%$$

Figure 7 – The WER model (Dumouchel *et al.*, 2011)

Nonetheless, the model focuses only on the quantity of words that are added, substituted, or omitted, thus not working effectively for cases when the text has been successfully reformulated or edited (Romero-Fresco & Martínez, 2015) – e.g., when filler words, question tags, repetitions are omitted in the TT, but the content-driven parts are delivered instead.

4.2.2 The NER model

The NER model (number of words, edition errors, recognition errors) was first introduced in Romero-Fresco (2011) with the NERD model (Number of words, Edition and Recognition errors, Deducted marks) (Romero-Fresco, 2011: 150-161), and then further developed in 2015 drawing from the WER model (Romero-Fresco & Martínez, 2015), this time highlighting the need for human intervention in assessing the quality of subtitles. The model was also applied to a study by Ofcom (UK Office of Communications) to measure quality in intralingual live subtitles (2014a, 2014b, 2015a, 2015b). The model accounts to N, which is the total number of words in the subtitles including punctuation; edition errors (E), and recognition errors (R), also accounting for successful editions that were made by the respeaker as correct editions (CE). The calculation consists of subtracting the sum of edition and recognition errors from the total number of words, divide by N again, and multiply by 100 to obtain the reference percentage. CEs are also counted and, though not mathematically included in the calculation, will have weight on the overall quality assessment.

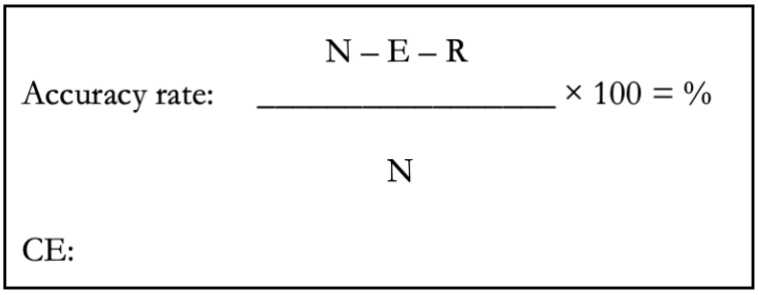


Figure 8 – The NER model (Romero-Fresco & Martínez, 2015)

Another interesting factor introduced by the NER model is that errors have different degrees of severity according to the impact they have on the TT, in that it focuses on meaning units called idea units rather than on single words, something which is also more human-centered given that respeaking and reformulating reference for accuracy cannot only account for the weighted number of words in a sentence or a text. Error grading is

classified as follows: minor errors (0.25, inconsequential deviations), standard errors (0.5, isolated information loss) and serious errors (1, utterance with a new meaning), and the expected threshold of the calculation of accuracy is 98%. The NER model and the types of errors it identified were applied and verified during the DTV4ALL project (Romero-Fresco, 2010)[13].

4.3 Assessing interlingual live subtitling

Moving from intralingual respeaking quality assessment to interlingual, several attempts were made to design a model that would be, above all, easy to apply to different language pairs and that took into account ST and TT and acknowledged that not all errors have the same impact. Soria (2016) proposed the NERT model in a first attempt to adapt the NER model for intralingual live subtitling to their interlingual declination. Soria: "[makes] a distinction between E (edition errors) and T (translation errors). It is difficult to see, however, how (interlingual) editing could be differentiated from changes in meaning or content" (Romero-Fresco & Pöchhacker, 2017: 160). More well-known and widely applied at an international level (Eugeni, 2017) is the NTR model by Pablo Romero-Fresco and Franz Pöchhacker, first published in 2017.

4.3.1 The NTR model

The NTR model (number of words, translation errors, recognition errors) was co-designed by the two professors and researchers in 2017. The model was first applied during PhD research on interlingual respeaking that began at the end of the same year (Dawson,

[13] https://cordis.europa.eu/project/id/224994/it.

2020), and then also applied to assess interlingual respeaking outputs for the SMART project experiments. The model can be applied to assess subtitles produced in any language (TT) and from any language (ST), thus the presented doctoral research heavily relies on this model since it was used to carry out analyses for both experiments, with English to Italian and Spanish to Italian outputs. The NTR model draws upon the NER model, proposing a similar formula and error grading process, but this time focusing on the interlingual shift from one language to another, thus accounting for translation errors instead of edition errors. Therefore, the calculation consists in subtracting the sum of translation errors (T) and recognition errors (R)from the total number of words in the subtitles (N), dividing by N again, and then multiplying by 100: the reference threshold rate for a subtitle in order to be deemed acceptable is 98%. Correct editions of the NER model, namely those that do not lead to a loss of information or that are considered strategic while respeaking interlingually, are now called 'effective editions' (EEs).

$$\text{Accuracy rate: } \frac{N - T - R}{N} \times 100 = \%$$

EEs:

Figure 9 – The NTR model (Romero-Fresco & Pöchhacker, 2017)

In this model, translation errors refer to the performance of the respeaker, while recognition errors refer to misrecognitions by the SR software. As Figure 9 below shows, translation errors are divided into error types based on content or form, the former accounting for omissions, additions and substitutions, the latter for style and correction errors. As in the NER model, error grading is threefold and is categorized using the following terminology: minor errors (0.25), major errors (0.5), critical errors (1).

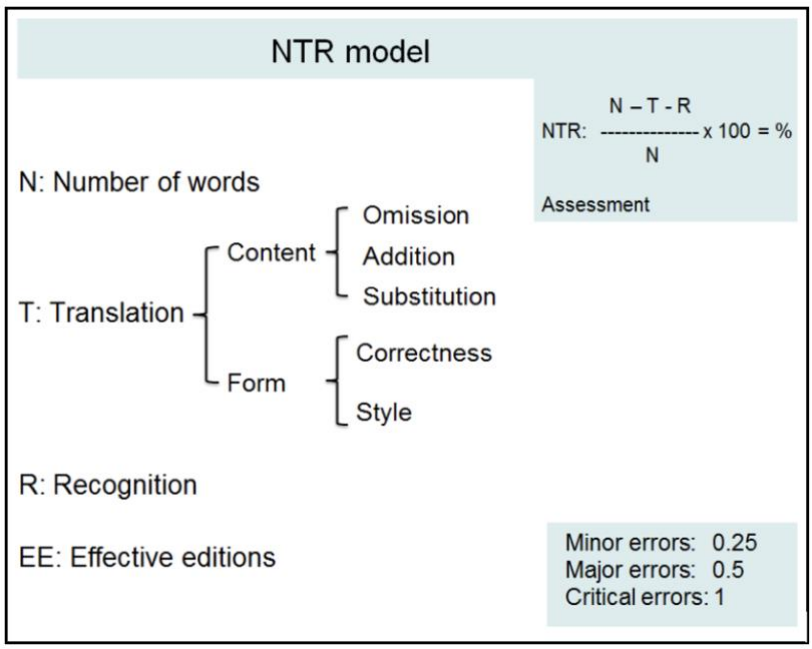


Figure 10 – The NTR model (II) with error type categorizations (Romero-Fresco & Pöchhacker, 2017)

As for the previous models, the NTR formula caters for linguistic accuracy calculation but the model also includes other factors in the overall quality assessment of the subtitles, such as the positive impact of effective editions on the TT, and the speed and the delay of the subtitles. As pointed out by the authors:

> "In the case of a discrepancy between the accuracy rate and the conclusions from the broader assessment (for example, when the delay of highly accurate subtitles is considered unacceptable), it is the final conclusion and not the accuracy rate that represents the quality of a given set of subtitles as assessed with the NTR model".
>
> (Romero-Fresco & Pöchhacker, 2017: 159)

In addition, the new model recognizes that high percentage rates are more common in intralingual subtitling given the complexity of ILS, it therefore recalculates the accuracy rate on a 10-point scale, with a reference percentage scale (Table 1). An interesting further classification of performance is provided by Dawson for more accurate assessment (Table 2). Rates for the analyses of this research refer to both the accuracy percentage and the relevant reference on the 10-point scale, and the classification proposed later.

Accuracy %	**10-point scale**
< 96	0/10
96.4	1/10
96.8	2/10
97.2	3/10
97.6	4/10
98.0	5/10
98.4	6/10
98.8	7/10
99.2	8/10
99.6	9/10
100	10/10

Table 1 – Ten-point scale calculation in NTR model (Romero-Fresco & Pöchhacker, 2017)

Accuracy %	10-point scale	**Classification**
< 96	0/10	Unclassified
96.4	1/10	Very poor
96.8	2/10	Poor
97.2	3/10	Poor
97.6	4/10	Satisfactory
98.0	5/10	Satisfactory
98.4	6/10	Good
98.8	7/10	Good
99.2	8/10	Very good
99.6	9/10	Excellent
100	10/10	Exceptional

Table 2 – Classification of performances in reference to the NTR model (Dawson, 2020)

4.3.2 Dependent and independent idea units

For the sake of both the NER and the NTR models it is important to make a difference between independent and dependent idea units. As previously mentioned, starting from the NER model there was a shift toward units of meaning rather than single weighted words. By units of meaning, we refer to independent idea units, which are usually full sentences since they are meaningful units containing a full message. On the contrary, a dependent idea unit specifies some aspect of the independent idea unit. It can be said that independent units are made up of several dependent units that provide information or add detail on the 'when', the 'where', the 'how' (Eppler & Ozón, 2013 in Romero-Fresco & Pöchhacker, 2017). Therefore, independent idea units make sense as a whole, while dependent idea units do not alone convey any meaningful information since they are complements and therefore optional parts of the sentence without which meaningful units are still provided. Error grading for edition (in the NER model) and omission errors (in the NTR), therefore not for recognition errors, derives from this distinction, suggesting that the omission or the edition of a dependent unit would have less burden on the TT than the

omission of an independent unit, even if this was not always the case concerning the type of specification that the dependent unit provides for the independent. We will now observe error grading in more detail as it forms a highly pertinent part of the analyses that provided data and results for this research.

4.3.3 Error grading

We will now focus on the error grading identified for the NTR model by Romero-Fresco and Pöchhacker (2017) which also resembles and draws upon the similar distinction for NER as well, with standard errors becoming 'major', and serious errors now becoming 'critical'.

Both for translation and recognition processes, minor errors cause only a small loss of content and do not impact comprehension or meaning. These minor errors are mainly recognition errors and mainly relate to punctuation, capital letters, and the mistaken addition (or misrecognition) of small words.
A major error implies a loss of content depriving the reader of an idea unit without them noticing. Major translation errors would then be represented by the omission of a full independent idea unit, while major recognition errors account to those that the reader can identify (either a misrecognition resulting in a nonsensical word that the reader recognizes as a mistake without guessing the ST meaning, or a sensical new word that can be clearly deemed a misrecognition since it makes no sense in the context).
Finally, critical errors introduce new content unrecognizable to the reader and constitute misleading information. Critical errors mainly occur during the translation process due to information substitution.

While recognition errors (R) do not have subcategories, minor, major and critical translation errors (T) can relate to both content and form.

- Content omission errors refer to errors caused by the omission (deletion) of some part of the ST. Omission of dependent units would normally account to minor errors, while omission of full independent units to major;

- Content addition errors refer to extra information in the TT that was not in the ST. Small additions that do not add extra content-driven parts would normally account to minor errors;

- Content substitution errors refer to errors caused by some information being substituted with a new idea, resulting in the introduction of new, misleading information;

- Form correctness errors refer to errors of grammar or terminology in the TT;

- Form style errors refer to errors caused by unnatural translation or changes in the register.

Lastly, effective editions (EE) can be omissions or substitutions (reformulation, generalization, etc.) that do not lead to loss of content. An EE is spotted when the ST has been modified to a certain extent but without losing relevant information.

4.4 Assessing SI (and respeaking)

With reference to the subject matter of Section 4.1 regarding the concept of quality (*ibid.*), it is important to stress how challenging assessing quality in ILS can be (Robert & Remael, 2017). Quality assessment is indeed complex per se in SI (and translation) (Gambier *et al.*, 1997; Pöchhacker, 2001; Errico, 2016), therefore by combining SI and SR techniques in respeaking, it seems fundamental to offer an overview of how evaluation of translated products has been historically attempted. Can translation and interpreting (and consequently respeaking) be assessed on a merely lexical level (counting the number of words that are omitted, added, or changed)? Can transfers between different languages be

compared based on linguistic equivalence? In the same light, can equivalence be considered at a syntactical level in two different languages? These are some of the questions that stimulated the analyses carried out for this research. To guide readers through the considerations that were drawn as conclusions (discussed in Chapter 4), an overview on some of the different parameters and positions for assessing SI processes is now given.

In Translation Studies, qualitative analysis of translation and interpreting products has always been challenging and debated (Pym, 2010) since it depends on many factors. Over the years, different approaches concerning the concept of 'translation equivalence' have developed.

A first approach is that of theorists who "seem to forget that translation in itself is not merely a matter of linguistics" (Leonardi, 2000: 1), but rather also a matter of linguistic relations in both SL and TL, including semantics (Romero-Fresco & Pöchhacker, 2017). These translation scholars would naturally support the 'translation equivalence' model, assuming there could be an exact correspondence between the SL and the TL without interference (Barik, 1975; Vinay & Darbelnet, 1995). Some scholars argue for different types or degrees of equivalence instead (Nida, 1964; Nida & Taber, 1969/2003) – formal equivalence and dynamic equivalence[14] – while others allow for paraphrasing in the translation process without strictly focusing on the number of words correctly translated, thus overcoming the mere lexical level of analysis in seeking a more meaning-based category of analysis (Gerver, 1969/2002, in Romero-Fresco & Pöchhacker, 2017: 156).

When approaching the concept of quality in interpreting, Viezzi (1999: 146-151), for example, defines it as the balance between four different factors, for example, that is: equivalence, accuracy, appropriateness, and usability[15], showing how different variables are

[14] Formal equivalence aims at achieving equivalence between the ST and the TT reflecting the same linguistical features as vocabulary, grammar, syntax, and structure, assuming the reader knows the cultural and linguistic context of the ST itself. The dynamic equivalence takes into account the reader and aims at transporting the message of the ST in the TT so that the receptor perceives it in the very same way of the original.

[15] Equivalenza, accuratezza, adeguatezza e fruibilità (Viezzi, 1999).

involved in the assessment of quality itself. Referring to the concept of 'equivalence' in quality assessment for translated and interpreted products seems difficult – if not impossible. In fact, when observing accuracy and equivalence as the correspondence of same content information in the TT, Viezzi points out:

> "Una certa quantità di perdita di informazione può essere sempre messa in conto nell'interpretazione e [...] non rappresenta necessariamente un fattore negativo ai fini della comunicazione e può essere inoltre considerata un sacrificio necessario nel perseguimento di altri obiettivi dell'interpretazione[16]".
>
> (Viezzi, 1999: 147-148)

As pointed out in Errico *et al.* (2021) – despite focusing on consecutive rather than simultaneous interpreting, and therefore on interactive and pragmatic level of communication that are less crucial in remote respeaking – omission as a deliberate act can be deemed strategic. In respeaking as well, omissions can be a strategic choice to keep up with the source speed, for example.

Opposed to these first views of the need for 'literal' translation, another approach was developed according to which a full equivalence between two different linguistic codes can not exist, thus shifting towards more pragmatic and semantic features of translation and interpreting (Jakobson, 1959) and again more meaning orientated.
The myth of equivalence in the translation process was gradually substituted this way by this more realistic perspective, arguing that:

[16] Some loss of information is expected in interpreting and […] is not necessarily negative for communicative intentions as it can also be seen as a sacrifice for the sake of other aims in the interpreting process (my translation).

> "consistency of meaning (or sense) between source and target texts (also referred to as fidelity, or accuracy and completeness) is the most highly valued criterion in assessing the quality of an interpretation […]".
>
> (Romero-Fresco & Pöchhacker, 2017: 157)

This more functional approach was implemented throughout the analyses for the experiment, temporarily giving up on the search for a 'perfect fit' quality assessment for SI outcomes. Particular emphasis was made on units of meaning rather than single words, or single dependent vs. independent unit analysis given the cognitive load that interpreters – and more so interlingual respeakers – must manage. In a nutshell, more importance was given to a taxonomy based on idea units, the ST content, and ideas being conveyed overall in the TT rather than at lexical or syntactical levels. Syntax and lexicon, after all, are difficult to compare depending on the language pairs, since formal grammar does not always correspond to meaning. This is especially true in interlingual translation. Take as an example Spanish and Italian, which have similar syntactical structures, while for English and Italian or other language pairs it can be more complex.

4.5 The issue of subjectivity in assessing SI (and respeaking)

After explaining some of the premises on the level of analysis concerning the TTs of the experiment, it is worth mentioning another very important issue: subjectivity. Subjectivity in scoring error severity during the analysis is a crucial point that must be dealt with since it can play a major part when it comes to assessing meaning idea units, rather than lexical or syntactical levels. In short: can we neatly establish which level of grading (error severity) to attribute to each error, and how? Depending on the context of the sentence and the interpretation carried out, it is possible that a range of opinion exists on

how much each specific error might impact meaning, content, and comprehension in the TT. A minor severity error for one could amount to a critical severity error for another, for example, or some loss of content could be considered strategic condensation for one (thus an effective edition), while for another could be considered a content omission (usually a minor severity error depending on how much content is omitted).

As highlighted by Romero-Fresco and Pöchhacker themselves:

> "The use of appropriate strategies and techniques, such as compression, generalization or stalling for time, is considered an important part of interpreters' professional skills, many of which are needed to cope with the high processing load associated with the task".
>
> (Romero-Fresco and Pöchhacker, 2017: 157)

Kalina (2015) notes that strategies such as reformulation and strategic compression (omissions) can indeed be considered effective strategies in interpreting, thus not always accounting for errors. In such cases, subjectivity plays a major role in labeling compression-omission, for example, as an effective edition, or as an error. On other occasions, even the omission of an independent idea unit – that would normally amount to a major error (-0.5 points) – could arguably instead amount to a minor omission. This is the case, for example, for different rhetorical stand-alone questions that expressed very similar ideas and that can reasonably be omitted by simultaneous interpreters or intra and interlingual respeakers for reasons of speed. As noted, it is of course impossible to determine whether the omission was strategic and left out deliberately, or whether it was due to the overload, the ST speed, or other factors (Gile, 2009; 2011), and therefore attributable to the respeaker/interpreter's weaker performance. Unfortunately, the model does not allow for evaluating such parameters that are, consequently, left unspecified. As Pym (2008) suggests, such omissions

are categorized according to their impact on communication: if the omission is deemed strategic, regardless of the awareness of the performers, it does not amount to an error.

As pointed out by the authors testing both the NER and the NTR models, one solution for minimizing subjectivity is to have more than one evaluator carrying out the assessment (inter-annotators), to monitor any discrepancies between different people's opinions. This process was applied to the present research as well, as will be detailed in Chapter 4 when dealing with the outputs analyses and results.

Chapter 3.

Interlingual respeaking workshops

1. Introduction

Chapter 3 will provide an overview of the interlingual respeaking workshops offered at the University of Genoa (UniGe), the aim of which was to train participants for the relevant experiments. After completing the training of the researcher (also an interpreter) through the ILSA course, the two introductory courses previously referred to as workshops were proposed as non-compulsory modules for the master's degree in Translation and Interpreting. Both training programs – the first being English to Italian in 2020/21 and the second Spanish to Italian in 2021/22 – followed the same structure: participants, modules and teaching methods, training materials. Nevertheless, there were some substantial differences between the first and the second edition mainly concerning the structure of the modules. The participants' information is displayed via answers to the pre- and post-workshop questionnaires, also detailing their knowledge in the field of MA and STTI. The chapter ends with answers to the post-workshop questionnaire, investigating students' acquired skills and level of satisfaction.

2. The training: interlingual respeaking workshops at UniGe

As previously mentioned, to accomplish the objective of assessing quality in different ILS methods and answer the main RQ for this research, two workshops in interlingual respeaking were designed to introduce conference interpreting master's

students to the technique for the first time. This training was offered to train those students that would also participate in the experiment phase. Drawing upon some of the training proposals for interlingual respeakers (see Chapter 2, Sections 2 and 3), two training workshops were offered at the University of Genoa with this purpose in mind.
The first ever workshop on intra and interlingual respeaking took place during academic year 2020/21 from English to Italian. Shortly after the completion of the ILSA project, materials were tailored and adapted for our students and online introductory training was proposed involving 15 students that would make up the test group for the main experiment of this research. The experiment tested three different methods of producing ILS – interlingual respeaking, SI + intralingual respeaking, SI + automatic SR.

A second interlingual respeaking workshop was offered the following academic year 2021/22, this time from Spanish to Italian, thus shifting scope and testing a differing language pair. In this last case, the workshop and the experiment were completed only a few months before the thesis was ready. Only 5 students were trained and participated in the testing during this round, but in this second experiment five methods to produce ILS were tested: those analyzed in the first round with the addition of two more machine-centered methods: intralingual respeaking and Machine Translation (MT), and ASR and MT.

Both training experiences are purposely named 'workshops' and not 'courses', since they cannot realistically be compared to a full training course of six-months, which would be the optimum for students in their early stages of learning SI and subtitling. Both workshops lasted approximately 70 hours each – including synchronous, at-distance learning and individual practical exercises – and were taught over a three-month period. Students subscribed voluntarily since the training was not an integral part of the master's degree, rather an introductory experience new to them which provided 3 CFU (ECTS) to their academic career. It is worth noting that during the first workshop, only some of the students required the relative ECTS; many did not need them and attended the workshop out of professional interest. Both training workshops were similarly structured although in the second, room for improvement at the end of the first educational and learning experience

highlighted some crucial areas that required further training. The structure of the second workshop was adapted accordingly as will be detailed in Section 4, as were the materials reflecting the different language pairing (English/Spanish). Pre- and post-workshop as well as and pre- and post-experiment questionnaires were submitted to the participants on both occasions to understand their backgrounds and levels of knowledge in the field of MA and respeaking, to monitor their learning curves and satisfaction, and also their self-assessment during the experiments.

We will now examine the structure of both workshops, starting with the first, delivered from English to Italian.

2.1 Directionality

The issue of directionality has long been debated in translation and conference interpreting studies, and for this reason it is also relevant to the field of respeaking. In choosing training materials for the interlingual respeaking workshops and the research experiment videos, the issue of directionality was considered.

Directionality itself relates to translation and interpretation that is carried out into or from a person's mother tongue. The expression 'direct translation' refers to translation into the translator's mother tongue/s, while 'inverse translation' (indirect) refers to translation from the mother tongue to a foreign language. The same applies for interpreting, namely direct or inverse interpreting. In Italian, this is what is usually meant by 'active languages', the ones translated or interpreted from and to (with *retour*), or 'passive languages', i.e., the ones translated or interpreted from. In translation and interpreting, one's primary language is usually identified as 'A language', while other working languages with B, or C and so on depending on how many are spoken. Normally, according to the *Association Internationale des Interprètes de Conférence* (International Association of Conference Interpreters, AIIC), language A is the direct language – thus the one that is translated from and to, the native language, while language B or C are inverse translation languages. Whether or not

interpreters should work towards their B and C languages is an old and debated issue (Gile, 2005). Some scholars, for example, do not support retour interpreting for the higher cognitive load it demands if compared to an interpretation into one's mother tongue (Seleskovitch & Lederer, 1989), still it seems difficult to generalize a situation that depends on many factors such as the individuals' level of experience and language knowledge, language pairs, and different standards in different countries. Nevertheless, some attempts were made to survey and compare direct and inverse translation in professionals. The PACTE group conducted a survey in 2004-2005 (Hurtado Albir, 2017) that highlighted how among the participants the trend was to translate into their mother tongue for several reasons. Quality of the translated products was higher in direct translations; those surveyed perceived inverse translation to be more difficult, while direct translation texts were deemed easier. Time taken to translate was shorter when direct, and the need to use external documentary resources occurred much more often during inverse translation. It was also found that translation from the dominant language to a weaker B or C language takes longer and causes longer reaction times than translating (or interpreting) in the other direction (de Bot, 2000).

Although a number of scholars support instead the bidirectionality of interpreting, for the interlingual respeaking workshops, as well as for the experiment material, it was decided to focus only on direct interpreting, therefore from the students' B language (for the first edition of the workshop English, and for the second Spanish) to their mother language which was Italian for all. Due to the already complex nature of interlingual respeaking, it was not our aim to further complicate the students' assignments. In addition, very little inverse translation or interpretation is normally carried out at the Department of Modern Languages and Culture of the University of Genoa, thus it would have made little sense to introduce it. For these reasons, the workshops were only offered from English and Spanish to Italian, and not vice versa. Furthermore, accent of non-mother tongue speakers may not be a relevant problem in inverse SI, but it could represent an obstacle in inverse ILS due to potentially more frequent misrecognitions. In fact, slightly mispronounced words or accents that sound more or less natural – either in English or in Spanish – would probably be easily understood by the audience, regardless being it native speakers in the TL, or L2

speakers. In ILS, instead, unnatural articulation of words or accents could result in more recognition errors by the machine set on the TL.

3. English to Italian interlingual respeaking workshop

The first workshop was delivered in 2020, starting in October and lasting approximately 70 hours throughout a three-month period until the end of January 2021 (twelve weeks training plus the experiment, developed in February 2021). The training was conducted online due to the Covid-19 pandemic and was delivered via Microsoft Teams, the university's official platform. The program was fairly dense as students attended remote two-hour lessons twice a week on average, and attendance for those students who voluntarily subscribed was monitored, since compulsory to obtain the ECTS. Synchronous lessons amounted to approximately 50 hours of training, while the remaining hours were dedicated to practical exercises and individual training. Firstly, a theoretical introduction was provided on MA, accessibility and respeaking technique to allow students to get up to speed on the subject matter of the workshop. An introduction to subtitling and a review of interpreting with preparatory exercises were provided before moving onto intralingual respeaking and focusing on SR software use. Lastly, interlingual respeaking practices (English to Italian) were covered. Modules and materials for the workshop will be more explicitly detailed in Section 3.2.

3.1 Participants

As mentioned, students who participated in the interlingual respeaking workshop also took part in the research experiment on completion of training, which was designed to be preparatory to the experiment itself. 15 students participated in total: 11 females and 4 males aged between 23 and 27 years old. They were all students enrolled in the master's degree in Translation and Interpreting of the University of Genoa (LM94): two students in

their first year, and the other 13 in their second years or about to graduate. Very few of them had previous experience in pre-recorded subtitling, while all of them had some experience in SI training, albeit from a variety of backgrounds. All participants specialized in English as their working language; most (9 of 15) worked with German as their second, the others being French, Spanish and Russian. They had previously obtained a bachelor's degree with the same working language pairs either in 'Language Mediation' (*Teorie e Tecniche della Mediazione Interlinguistica*, L-12), 'Modern Languages and Cultures' or the former 'Foreign Languages and Literatures' (*Lingue e Culture Moderne*, *Lingue e letterature straniere,* L-11). Only one participant had not graduated with English as part of their studies, however, already spoke it fluently. Therefore, all the participants had good levels of English, despite this not being measurable through official certificates or the Common European Framework of Reference for Languages (CEFR). According to the Italian Ministry of Education (MIUR, Ministero dell'Istruzione, dell'Università e della Ricerca) students at their level should have a level around B2+ and C1[17].

3.1.1 Pre-workshop questionnaire and activity

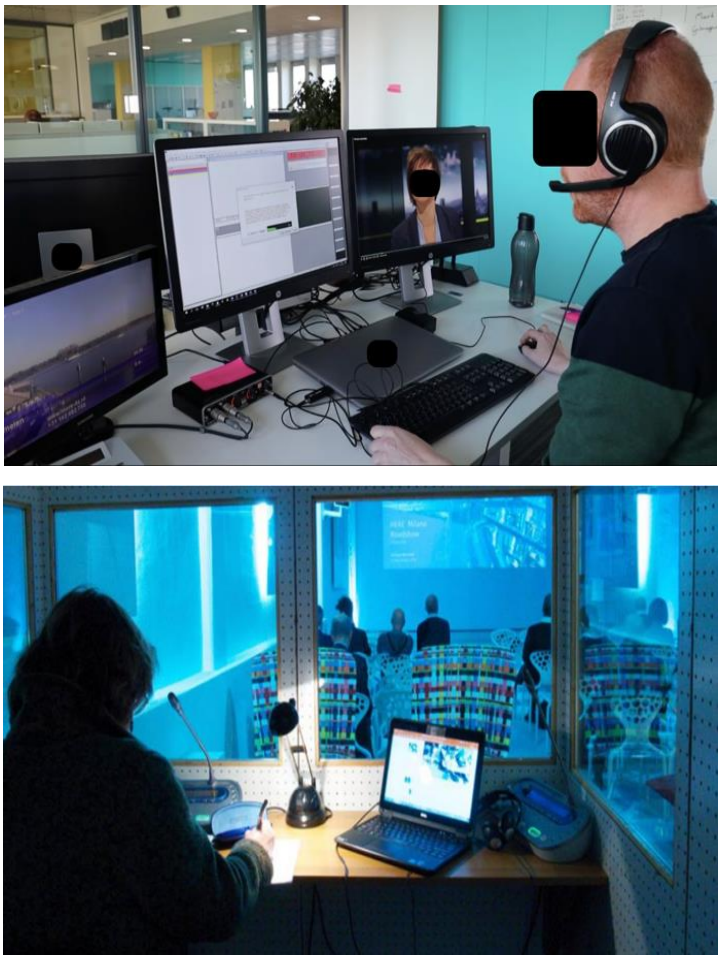
Background information on the participants was received via both a Wooclap activity[18] and a preliminary pre-workshop questionnaire using Google Forms.

The Wooclap activity was proposed to provide a more informal method of information and background sharing between the students, as many of them did not previously know each other, and to spark some preliminary discussion on the topic. The activity was carried out before the training started; the participants were asked if they knew any respeaking

[17] Art. 4, paragraph 1 of Italian Ministerial Decree of 7th March 2012 states that a two-years masters' degree ("Laurea Magistrale") in foreign languages corresponds to the C1 level (expert) of the CEFR, but no clear reference is made about the bachelor's degree instead. In any case, the degree does not correspond to an official language certificate, but are deemed to be only indicative.

[18] Wooclap (https://www.wooclap.com) is an online platform for creating surveys and questionnaires, among other ludic interactive activities. It aims to encourage students to be more active in their learning.

techniques and, if yes, where and how had they approached the topic. They were also presented with two pictures showing apparently similar working settings: a simultaneous interpreter in a soundproof booth with headphones, a laptop, and a microphone; and another professional (a respeaker in this case) wearing headphones and using a microphone, in a room with three screens, one of which displaying a piece of text, and one alphanumeric keyboard (Figures 11 and 12 below).



Figures 11 and 12 – A respeaker (above), and a simultaneous interpreter (underneath)

This input aimed at detecting the participants' perceptions between the two settings before they were trained in respeaking. Almost all of the students responded believing that both pictures showed simultaneous interpreters, some that the second professional was a subtitler, and only few reached the conclusion that the second picture could be a respeaking

task, but they were not sure about it. These answers preliminarily showed that the participants had very little or no awareness of the existence of another technique with a similar working environment and tools (see Section 3.1.3 below for further considerations about participants' knowledge on MA and respeaking). Finally, they were asked to brainstorm ideas and topics related to SI and, despite not being adequately prepared, they were asked to identify and brainstorm some respeaking features as well, this time highlighting some of the more specific features pertaining to respeaking such as 'media', 'accessibility', 'real-time subtitling', 'non-hearing people', and 'usability'. None of them matched with the brainstormed SI features, which only focused on the process of the task ('booth', 'anxiety', 'experience', 'speed, haste and concentration').

The Google Form questionnaire was submitted to gather some personal information about the students, and it is available to view in Appendix 2. Questions concerned each participant's gender, age, bachelor's studies field, and working languages. It was also useful to collect information about any previous training, covering their academic or professional experience/knowledge in the fields of:

- pre-recorded subtitling and/or SDH
- SI training in English, and also in other languages

Some other questions concerning the field of MA and, more specifically, STTI and respeaking were asked on completion of the training, also about their level of knowledge before the workshops.

3.1.2 Participants' background knowledge in subtitling and SI

Concerning pre-recorded subtitling or SDH, thirteen students had no experience in subtitling, either intralingual for DHOH or interlingual, while one participant had taken

(and another had previously worked on) a bachelor's degree introductory course on interlingual pre-recorded subtitling.

On the topic of interpreting, the students all had previous SI training either in English or in other languages but to different extents: some students from the second year of their master's had up to three years' training experience in simultaneous and other types of interpreting, some up to one year of experience, and two others had only recently approached the technique and had little training experience by then. Those who had more SI experience could rely on their training in split attention and control in interlingual production and were therefore seemingly more advantaged than those who had had less training and, thus, more likely to obtain a higher quality assessment. This is something that will be discussed in more depth in the results at Chapter 5.

3.1.3 Participants' background knowledge in MA, STTI and respeaking

Concerning the broader field of MA, as can be observed in Figure 13, participants were asked to rate their knowledge about MA before the workshop on a Likert scale from 1 to 5 (1 'strongly disagree', 2 'disagree', 3 'neither agree nor disagree', 4 'agree', 5 'strongly agree').

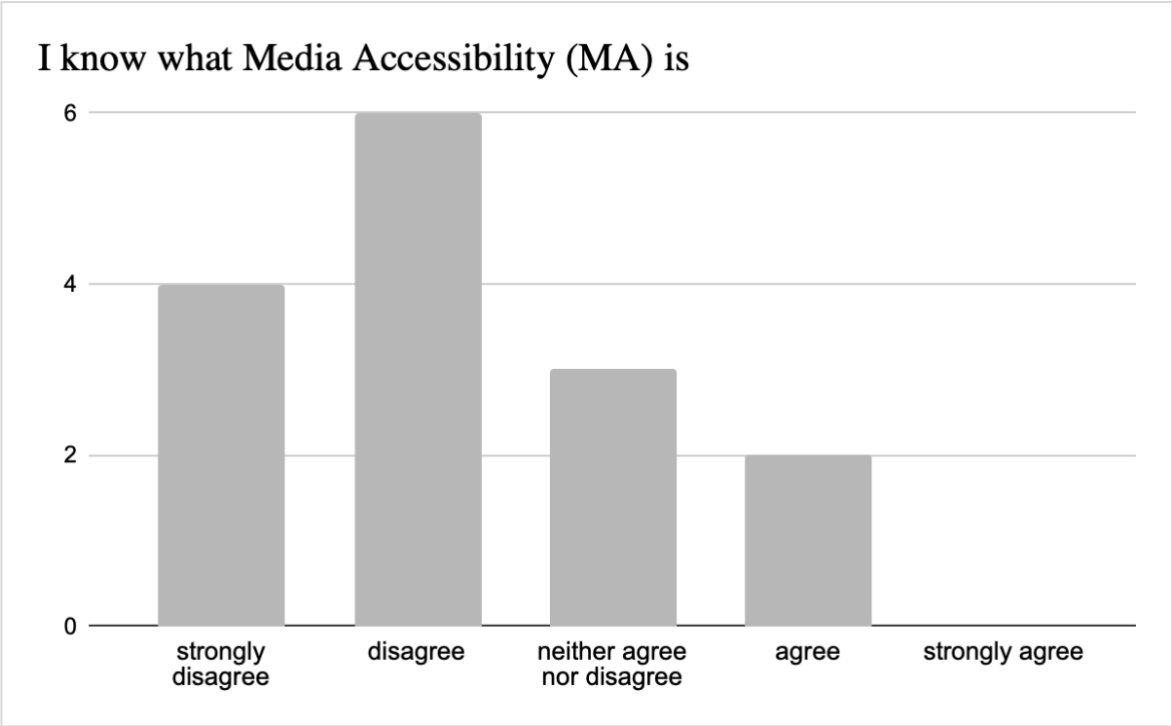


Figure 13 – Students' knowledge in the field of MA before the EN>IT workshop

Ten students declared that they did not know what it was, three that they knew – or could imagine – what it was, and only two students were confident enough to agree.

With respeaking, as can be observed in Figure 14 below, the students were asked if they knew of the concept before the training workshop. Again, ten students had never heard of it, four replied that they had heard of it but did not know exactly what it was, and only one had any previous knowledge of it.

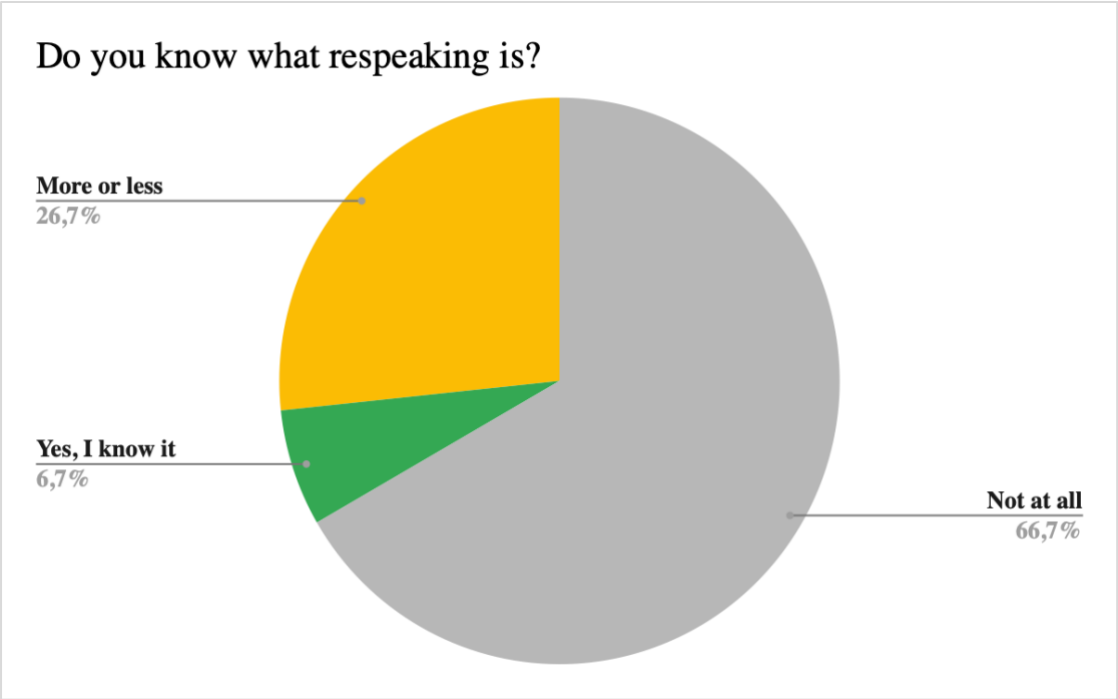


Figure 14 – Students' knowledge of respeaking before the EN>IT workshop

It was also pointed out that the students had subscribed to the workshop either because they did not know what respeaking was or had just heard of it and wanted to find out more about it – and one person because they already knew about it but wanted to receive formal training for professional reasons. These answers, as said, preliminarily showed that knowledge in the field was scarce among the students since almost all of them knew little or nothing about it, suggesting that adequate training in MA was both needed and urgent.

3.2 Teaching methods and training modules

Some of the materials for the workshop were taken from the ILSA course and tailored according to the students' levels of knowledge and needs. In particular, video lessons and reading tasks (as presented in detail below) were taken from the Campus do Mar MOOC platform by the Universidade de Vigo, where all materials for the interlingual respeaking course were open access and free. As shown in Figure 15, the first part of the

training was asynchronous, meaning that the materials were proposed to students so that they could access them freely and without time constraint. First, a theoretical introduction to accessibility in live events was provided, and then both theory and exercises on pre-recorded subtitling as the first proper module of the workshop. Synchronous lessons started with a debate on what students had read and were confronted with in the introductory phase. Module 2 focused on reviewing and strengthening SI, and module 3 on introducing intralingual respeaking through the SR software. Lastly, interlingual respeaking practices and contextual background were tackled, with the experiment signaling the end of the training path.

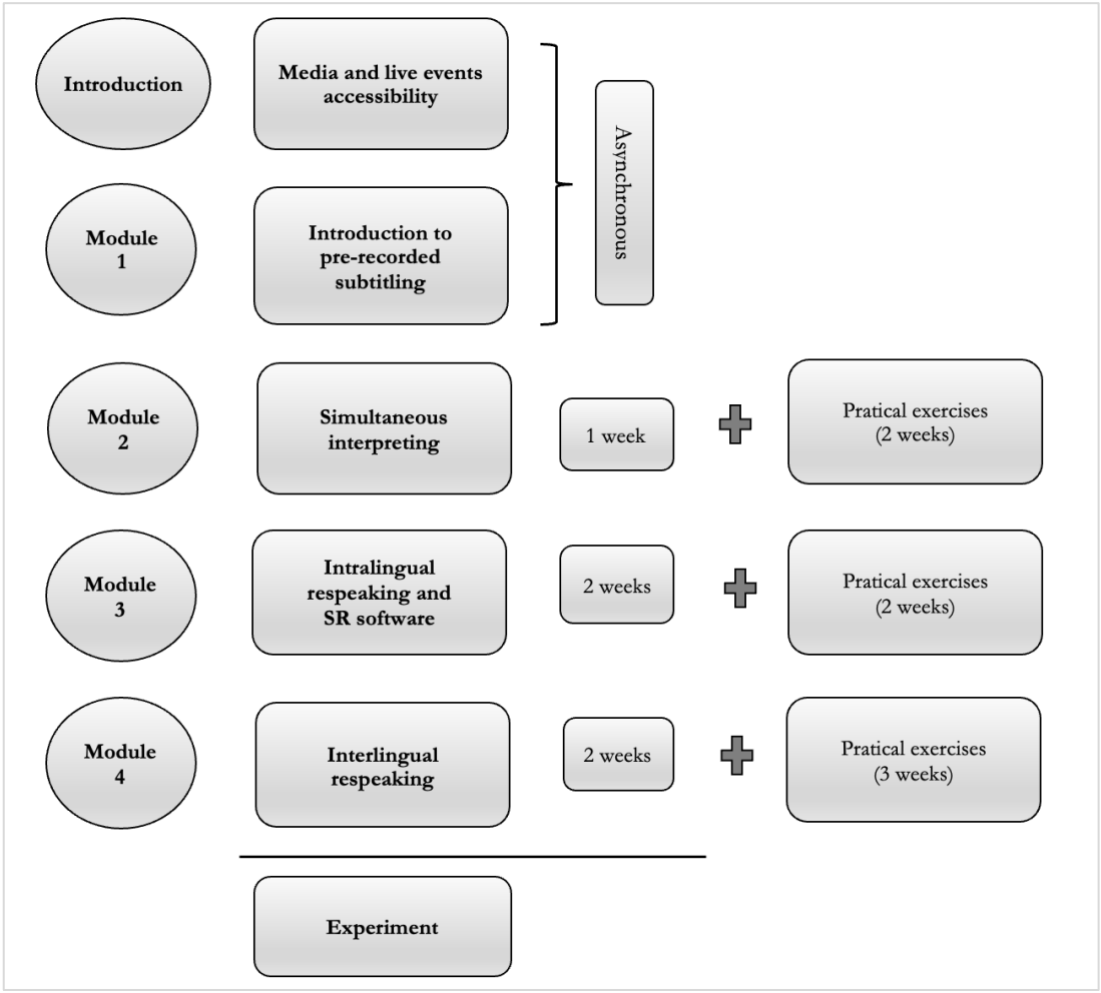


Figure 15 – An overview of the EN > IT interlingual respeaking workshop

Topics according to the modules were first introduced with an explanation of the tasks, competences and tools needed, as well as a process overview for each. After covering the theory, students moved to practice and approached the different tasks through set exercises. The resources varied from reading tasks to video lessons and tutorials or exercises. In detailing the different training materials for the workshop here, the different types of activities proposed to the students are signaled with different colors. Each section of each unit was either covered autonomously by the students or watched together with the rest of the class, although in both cases each topic was debriefed with the rest of the class. Therefore, for those materials labeled as video lectures, tutorials or reading tasks, note that this does not signify only passive learning by the students, but active through explanation and discussion of each activity in the classroom environment.

3.2.1 Media and live events accessibility

The workshop started with asynchronous lessons accessible to the students prior to the synchronous online lessons. Since many of the participants' main drive in subscribing to the workshop was to deepen their knowledge of accessibility, they were first introduced to the broader topic of disability and, consequently, on the need for accessibility in media through videos and short readings. The focus was then moved towards AVT, MA and live events, also introducing the field LS with particular attention paid to the technique of respeaking.

Introduction – Media and live events accessibility	
Unit 1: Human diversity and disability	
1.1 Human diversity	Reading task
1.2 Disability and accessibility	Video
1.3 Media accessibility and access services	Video
1.4 Media accessibility and Audiovisual translation	Discussion

Table 3 – Structure of the introductory section "Media and live events accessibility" of the workshop

3.2.2 Introduction to pre-recorded subtitling

The module in subtitling was used to introduce the difference between pre-recorded, semi and live subtitling and to dwell on features of subtitling practice that are pertinent to respeaking. After introducing the broader topic of MA and respeaking, this was the first and only module that students dealt with remotely by themselves.

A theoretical introduction to subtitling was given, highlighting typical characteristics such as condensation due to time-constraint, line breaks and subtitle duration between shot changes, for example. Despite the lack of consistent guidelines among different countries, prompts on maximum number of words in subtitles, on how to use labels to identify different speakers, and how to include music, sound and noise information in the subtitles were introduced. Students were also provided with a tutorial on how to use the EZTitles subtitling software with a demo on how to spot and time subtitles. They were also introduced to Aegisub, another type of free subtitling software, and attempted to subtitle short videos on their own.

MODULE 1 – Pre-recorded subtitling	
Unit 1: Introduction to subtitling	
1.1 Introduction to subtitling	Video lecture
1.2 Fundamentals of pre-recorded subtitling	Video lecture
1.3 Text condensation and line breaks in subtitling	Reading task
Unit 2: Learning to subtitle	
2.1 Subtitling guidelines and shot changes	Reading task
2.2 Subtitling software	Video tutorial
Unit 3: Subtitling for the deaf and hard of hearing	
3.1 Subtitling for the DHOH	Video lecture
3.2 Subtitling for the DHOH (II)	Reading task

Table 4 – Structure of Module 1 "Pre-recorded subtitling" of the workshop

3.2.3 Simultaneous interpreting

All students participating in the workshop had previous experience in SI to varying extents. Their performing level was also understandably varied, but they had all previously become familiar with the task, not only in English but also in other working languages. However, they retraced the different stages and focused on some preparatory and strengthening exercises for SI such as counting, shadowing, paraphrasing and summarizing, and sight translation (Lambert, 1989; Kurz, 1992). 'Counting' consists of specifically training the skill of split attention, which allows interpreters (and respeakers) to speak whilst continuing to listen to and understand what is being said – in other words, multitasking. Propaedeutic exercises of counting in reverse (e.g., from 100 to 1) while listening to an audio input were proposed. Intralingual shadowing calls upon a very similar task to respeaking, being "a paced, auditory tracking task which involves the immediate vocalization of auditorily presented stimuli, i.e. word-for-word repetition, *in the same language*, parrot-style, of a message presented through headphones" (Lambert 1992: 17).

This kind of exercises also helped students in training their *décalage*, listening and speaking at the same time, while paraphrasing consisted in a kind of shadowing technique that, without rendering verbatim the input, had them attempt to summarize the ST whilst looking for synonyms and similar expressions to convey the meaning. Lastly, interlingual practice was introduced through sight translation. Different texts were given to the students in English to translate out loud to Italian whilst maintaining consistent pace. Finally, the students had the opportunity to practice SI. To begin, speeches with a lower speech rate were given by the trainer, either with live speaking or recorded inputs. Gradually, faster and more complex real-time recorded videos were introduced. All materials used were general in their nature and contained no specific terminology. Videos were taken either from the Speech Repository for interpretation by the European Commission[19], where different level and domains can be selected for real-life or pedagogical material, or from TED talks[20], depending on their level.

It is also worth noting that, while attending the workshop, the students all continued to attend their parallel SI lessons in other languages practicing sight translation, as well as paraphrasing on their own.

MODULE 2 – Simultaneous interpreting	
Unit 1: Review of the technique of interpreting	
1.1 What is SI and its stages	Discussion
1.2 Interpreting and respeaking	Video lecture
Unit 2: Preparing for interpreting	
2.1 Guidance on shadowing	Reading task
2.2 Exercises on shadowing	Exercise
2.3 Guidance on paraphrasing and summarizing	Reading task

19 https://webgate.ec.europa.eu/sr/search-speeches?language=114&level=9&use=1&domain=All&type=All&combine=&combine_1=&video_reference=&entity%5B0%5D=

20 https://www.ted.com

2.4 Exercises on paraphrasing and summarizing	Exercise
2.5 Guidance on sight translation	Discussion
2.6 Exercises on sight translation	Exercise
2.5 Exercises on SI	Exercise

Table 5 – Structure of Module 2 "Simultaneous interpreting" of the workshop

3.2.4 Intralingual respeaking

After covering the introduction to subtitling and having completed the module on SI review, intralingual respeaking technique was approached. Different methods of live subtitling were shown, with special attention paid to respeaking, and the SR engine was introduced for the first time. Particular emphasis was placed on how recognition works in SR software, highlighting the need for proper segmentation when dictating, on how to dictate punctuation and words to the machine, and how to use the keyboard to correct any misrecognitions (or mis-dictations). Although they had the choice of both, almost all the students immediately found making corrections manually with their keyboard more convenient than using voice commands. Particular attention was then given to the software purchased by the Department for the experiment, 'Dragon Naturally Speaking' by Nuance. Professional individual licenses were purchased for the workshop to be installed in the students' computers so they could work from home and create their own user profiles.

MODULE 3 – Intralingual respeaking and SR software	
Unit 1: Live subtitling and respeaking	
1.1 Live subtitling and respeaking	Video lecture
1.2 Live subtitling methods	Discussion
1.3 Live subtitling methods	Video

1.4 Respeaking in Italy	Video
Unit 2: SR and dictation	
2.1 Speech recognition and dictation	Video lecture
2.2 Misrecognitions	Reading task
2.3 General dictation: punctuation	Reading task
2.4 General dictation: word stress	Reading task
2.5 Preliminary exercises	Exercise
Unit 3: Dragon Naturally Speaking	
3.1 Dragon Professional Individual – Administrator Guide	Guidelines
3.2 Dragon Instructions and tutorials	Guidelines
3.3 Dragon – User guide	Guidelines
3.4 Customizing the SR software	Video tutorial
3.5 Making live error corrections	Reading task
3.6 Preliminary exercises with SR software	Exercise
Unit 4: Intralingual respeaking	
4.1 Initial intralingual respeaking	Video lecture
4.2 Split attention and rhythm in respeaking	Reading task
4.3 Examples of initial respeaking by students	Videos
4.4 Exercises in intralingual respeaking	Exercise

Table 6 – Structure of Module 3 "Intralingual respeaking" of the workshop

Preliminary exercises consisted of intralingual shadowing exercises, encouraging enunciating punctuation out loud in order to get used to it while respeaking. Then, further preliminary exercises using Dragon were carried out, transcribing the recognized voice input on the DragonPad, which resembles a Word document, as shown in Figure 16 below.

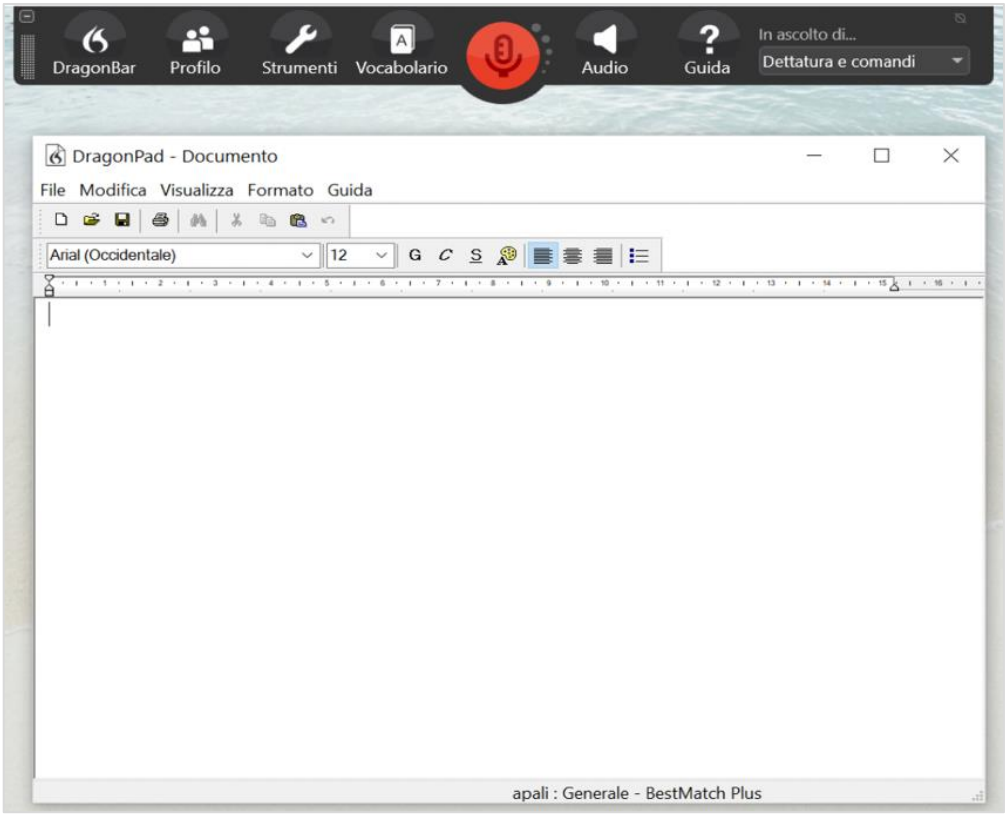


Figure 16 – Visualization of DragonBar for commands and DragonPad for transcription

3.2.4.1 Training materials

For both the SI practice and the intralingual respeaking exercises, generic subject matter was used that did not require any extra preparation on any specific terminology by the students. In SI, the speech rate variable considered comfortable in order to keep up with the ST is 100-130 words per minute (wpm), while 150-170 wpm is the threshold beyond which SI cannot be performed (Riccardi, 2015). Supposing that the same can be applied to respeaking, exercises were calibrated on gradually increasing speech rates (however, the added capacity required by respeakers to carry out additional tasks such as dictation to software, and even more so when shifting from one language to another, cannot be understated). First, some off-the-cuff speeches with a fairly slow speech rate (85-95 wpm) were given by the tutor to allow students to properly adjust to the task: listening while speaking into the software and trying to summarize subject matter when possible,

dictating punctuation marks and monitoring the transcribed output on the DragonPad in search of any misrecognition. Later, some faster pre-recorded speech (100-120 wpm) was introduced, while in the last phase even faster speech (up to 140-150 wpm) also entailed the additional complexity of slides being shown by the speaker on video, thus requiring as little delay as possible to synchronize as much as what was being said with what was being displayed on screen. These last practices were deemed to be particularly fit by the tutor for reformulation and condensation in order to keep up with the pace of the ST. The speech material used also varied in terms of duration: first students worked with shorter videos, and then videos of up to almost 20 minutes were respoken. Some of these pre-recorded real-life videos used as training material are detailed below.

Title	**Duration**	**Number of words**	**Words per minute (wpm)**
"L'amore e la felicità" da "I dieci comandamenti", Roberto Benigni	00:04:52	573	118
Part of "Meditazione 3.0", TEDx talk	until 00:05:11	628	124
"Chi è Salvatore Aranzulla", TEDx talk	until 00:05:04	703	140
Part of "Cosa vuol dire 'vero' su Instagram?", TEDx talk	until 00:04:13	586	144
Discorso di Liliana Segre al Parlamento europeo	00:19:07	1,645	85

Table 7 – Training materials for intralingual respeaking practice

3.2.5 Interlingual respeaking

Module 4 was the last module of the workshop and focused on interlingual respeaking. To prepare for the task, firstly the previous modules were reviewed, and before the practice began the stages of the interlingual respeaking process were analyzed. During the pre-process, emphasis was placed on the terminological preparation before the task, especially the addition and training of word pronunciation using Dragon. For some of the tasks, the students were provided with specific terminology or the generic theme to be addressed so they could create relevant glossaries and prepare in advance. Concerning software optimization, the post-process stage was also underlined as an important part of the debriefing process in which metacognitive reflection on performance is useful in highlighting any positives and negatives and identifying any new terms that can be added to the software, among other best practices.

MODULE 4 – Interlingual respeaking	
Unit 1: Introduction to interlingual respeaking	
1.1 Concepts and challenges	Video lecture
1.2 Bringing together subtitling, SI and intralingual respeaking	Video lecture Discussion
Unit 2: Stages of the interlingual respeaking process	
2.1 Stages of the interlingual respeaking process	
2.2 Pre-process	Video lecture Discussion
2.3 Peri-process	Video lecture Discussion
2.4 Post-process	Video lecture Discussion
2.5 Making live corrections working alone	Video lecture
Unit 3: Interlingual respeaking practice	

3.1 Sight translation (with punctuation) exercises	Exercise
3.2 Interlingual respeaking exercises	Exercise

Table 8 – Structure of Module 4 "Interlingual respeaking" of the workshop

Preliminary exercises to become familiar with the interlingual declination of the task were proposed through sight translation from English to Italian, in which students were asked to enunciate punctuation marks out loud whilst translating, imitating the process of dictating to Dragon.

3.2.5.1 Training materials

As per the SI and intralingual practice, the videos chosen for training during the interlingual respeaking module were generic in their subject matter and without any specific terminology. Similarly to the SI lessons, if any terminology was deemed too difficult or obscure due to the specificity of the subject matter i.e. acronyms and proper names, the students had prior sight of these in order to find the appropriate translation and input it into Dragon in advance. The same approach of gradually increasing the speech rate during intralingual respeaking was also applied for this module. First, pre-recorded speeches in English with a lower speech rate were given (approximately 85-100 wpm: former US President Obama's 'back to school' speech, Albert Einstein's Nobel dinner remarks, and former US President Trump's speech on the Mexican wall). Then, the students worked on pedagogical material videos from the Speech Repository at basic and beginner level at first, then intermediate, and finally various TED talks that were faster (approximately 130-160 wpm).

For the faster speeches, students were allowed to listen to the videoclips once before respeaking them, to adjust to the speed and note down any unfamiliar terms and better understand the context and topic. This was deemed important to reduce anxiety (Davitti &

Sandrelli, 2020) being respeaking such a complex task, especially for students that were not professionals in SI and this being the very first time they had attempted it.

Some of the training materials detailed below were pre-recorded videos; there were also recordings and ad-lib (or almost ad-lib) comments and reviews of movies or pieces of art, for example, given by the tutor.

Title	Duration	Number of words	Words per minute (wpm)
Unusual art performance "In orbit" description	00:07:46	952	123
"The future of work" Speech Repository	00:06:22	765	120
"The Lord of the Rings Trilogy overview"	00:09:50	1,167	118
"Friday 13th" Speech Repository	00:07:14	943	130
"The food app revolution" Speech Repository	00:05:49	656	129
"The origins of language" Speech Repository	00:09:11	1,294	141
"Mobile phones" Speech Repository	00:09:26	1,556	166
"The secret of becoming mentally strong", TEDx talk	00:15:00	2,374	158

Table 9 – Training materials for interlingual respeaking practice

18.3 **Post-workshop questionnaires**

At the end of the workshop, the participants were asked to complete two anonymous questionnaires via Google Forms. One was based on satisfaction and aimed at gathering feedback and thoughts on the recently completed training experience, the second was to monitor students' progression and preparation in the field of MA and respeaking. The first aimed at collecting feedback on the training in its different aspects, asking specifically about levels of satisfaction regarding the overall training experience and on teaching methods, setting, materials used, and acquired skills. Since it was the very first time that an introductory course on intra or interlingual respeaking had been provided at the university, it was deemed particularly relevant to collect the students' impressions and/or suggestions. Albeit anonymous, answers to the questionnaires can be partially conditioned by the awareness of students that their tutor would receive the feedback, thus they felt not entirely free to comment on downsides or negative aspects. The questionnaire comprised 13 questions. Some were open-ended, some were multiple choice or Likert scales ranging from 1 to 5. Concerning the course material, questions were asked on the balance between theory and practice sessions, any obstacles faced working remotely, the software and tools used, and finally the workload. Some questions also sought to find out about the future intentions of the students in continuing their training in interlingual respeaking, and whether they would advise colleagues and classmates to complete similar training. If the answer was positive, they were asked to briefly comment on why. Some very interesting and eloquent responses were received as shown in Figures 17, 18 and 19 below (the answers are given in Italian as this is how they were originally expressed, however a translation carried out by the researcher is available in the footnotes below).

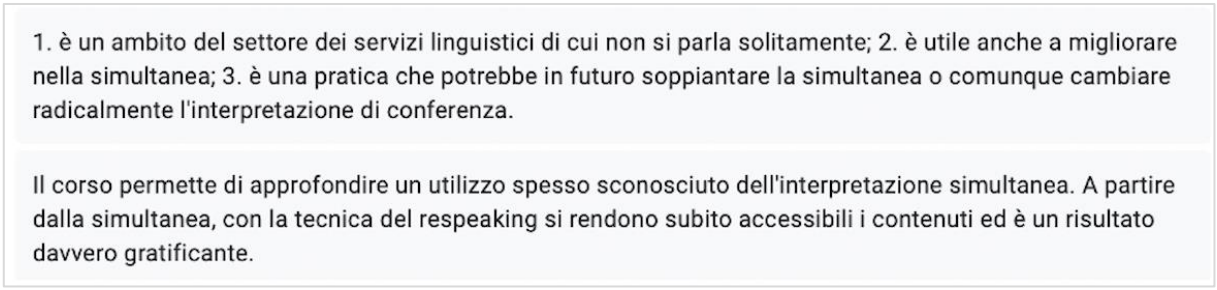


Figure 17 – Students' comments and opinions on the interlingual respeaking workshop (I)[21]

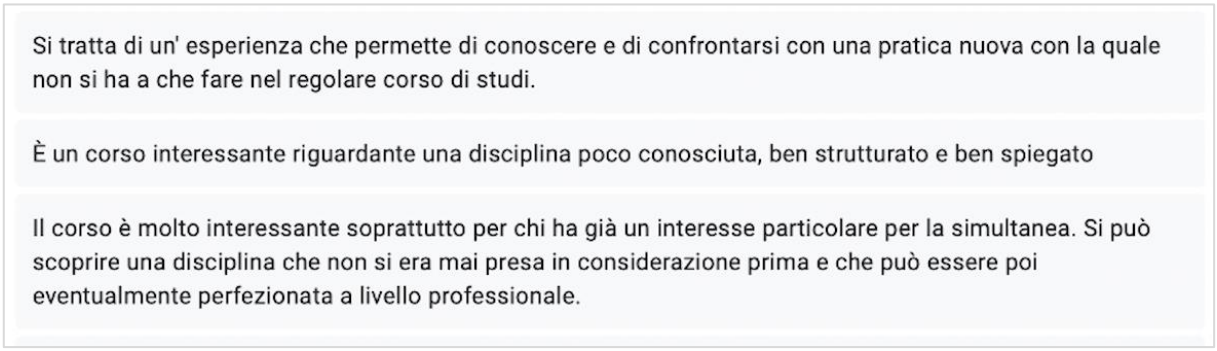


Figure 18 – Students' comments and opinions on the interlingual respeaking workshop (II)[22]

Something that was shared almost unanimously was the novelty of the training proposal as none of them had been confronted with something alike before. All trainees noticed the

[21] 1. It is a relatively unknown branch of the linguistic services field; 2. It is useful for strengthening SI; 3. In the future it could supersede SI, or at least radically change conference interpreting as we know it.

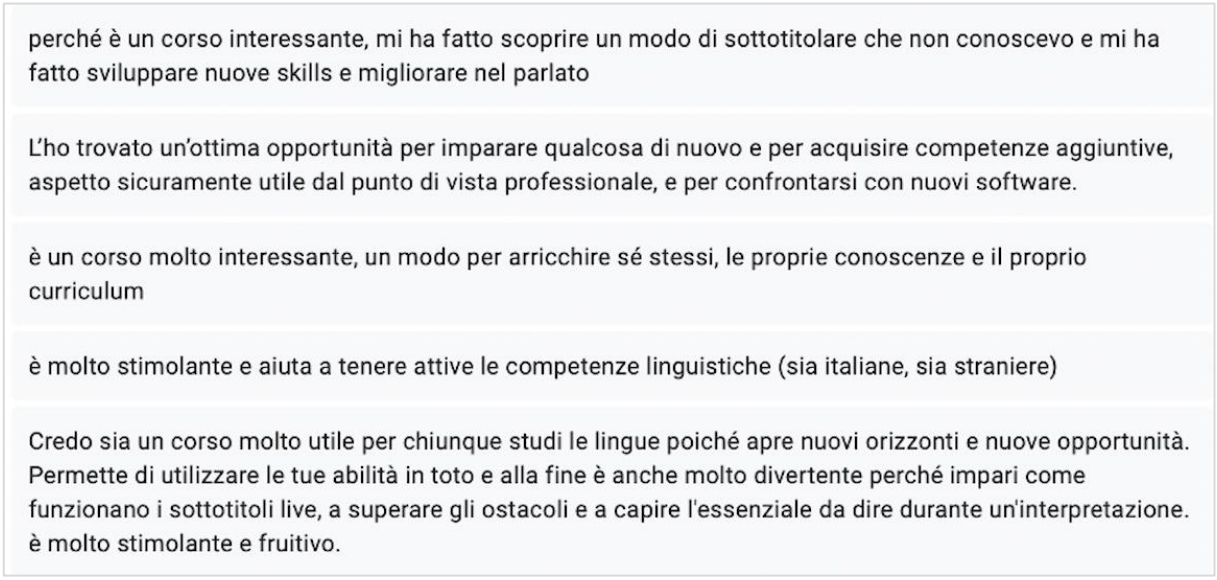
The workshop allows for deepening knowledge of a relatively unknown use of SI [, respeaking, through which] content is made accessible. I do think this is a very gratifying task at a personal level.

[22] This experience offers the opportunity of learning and being confronted with a new practice that is not included in the standard master's course [of our university].

It is an interesting workshop about an unknown discipline. It is well structured and well taught.

The training is particularly interesting for those who already are interested in SI. [Through it,] you can discover a discipline that has never been approached before [at our university], and that can be perfected for professional use in the future.

similarity between SI and respeaking, something that helped them improve their performance in interpreting itself, by using linguistic skills.


perché è un corso interessante, mi ha fatto scoprire un modo di sottotitolare che non conoscevo e mi ha fatto sviluppare nuove skills e migliorare nel parlato

L'ho trovato un'ottima opportunità per imparare qualcosa di nuovo e per acquisire competenze aggiuntive, aspetto sicuramente utile dal punto di vista professionale, e per confrontarsi con nuovi software.

è un corso molto interessante, un modo per arricchire sé stessi, le proprie conoscenze e il proprio curriculum

è molto stimolante e aiuta a tenere attive le competenze linguistiche (sia italiane, sia straniere)

Credo sia un corso molto utile per chiunque studi le lingue poiché apre nuovi orizzonti e nuove opportunità. Permette di utilizzare le tue abilità in toto e alla fine è anche molto divertente perché impari come funzionano i sottotitoli live, a superare gli ostacoli e a capire l'essenziale da dire durante un'interpretazione. è molto stimolante e fruitivo.

Figure 19 – Students' comments and opinions on the interlingual respeaking workshop (III)[23]

Students were also asked about any competences they felt they had acquired or strengthened thanks to the modules they completed. Many pointed out improvements in

[23] It is an insightful workshop that introduced a method of subtitling that I did not know about. It allowed me to develop new skills and improve my speaking competences.

I find it an amazing opportunity to learn something new and to acquire new skills, something particularly useful from a professional perspective. [It is also useful to] learn how to use new software.

It is a very interesting workshop, a way to enrich oneself, to expand one's knowledge and résumé.

It is stimulating and it helps to keep active one's linguistic competence in both languages.

I believe it is a useful workshop for anyone who wants to study languages since it widens horizons and offers new opportunities. It allows one to use their abilities in full and it is also quite enjoyable since you learn how live subtitles work, you learn how to deal with obstacles and to grasp the essential meaning to translate during an interpreting task. It is very stimulating and fruitful.

their SI technique thanks to the several skills it shares with respeaking. Among them, students quite interestingly pointed out all of the competences highlighted in Franz Pöchhacker and Aline Remael's study (2019), as well as Hayley Dawson's investigation (2020) in the search for sub-competences and sub-skills starting from the Effort Models (Gile, 2015). Here are the skills and competences that students felt they acquired newly, or improved through training: split attention, working under pressure, multitasking, problem solving. Without being told to identify them in the three different stages of the interlingual respeaking process, students automatically retraced the pre-, peri-, and post-processes through their acquired skills. All participants highlighted the importance of proper knowledge of how to use the SR software and how to prepare it by adding and training new words. Regarding linguistic competence (and use of adequate register) they highlighted listening comprehension, information management, strategic reformulation, and reduction as important; and regarding dictation they highlighted monitoring, editing, and good intonation. Very interestingly, reduction in reaction time was underlined as a new acquired skill. Concerning the use of the SR software, one participant notably pointed out that it was fundamental to "try to create from the beginning an output that is as clear as possible, in order to intervene as little as possible on the TT[24]".

It 117 salso relevant to highlight an answer where a participant pointed out: "Sicuramente ho migliorato le mie competenze da interprete perché ho capito l'essenziale che deve passare a un pubblico, senza 'aver paura' di dire troppo poco[25]". This comment seems to suggest that in both respeaking and interpreting it is of the utmost importance to interpret and respeak only what is essential, without lots of other information, which is of course not the case. What is believed the participant really wanted to express is regarding the need in interlingual respeaking to condense some information in order to keep up with the multitude of tasks and the ST pace, and that sometimes 'less is more' (in other words, it is

[24] "Attenzione nel generare un output il più chiaro possibile già da subito, per poi fare meno modifiche" (my translation).

[25] "I undoubtedly improved my SI skills since I understood which essential parts must be transmitted to the audience, without being 'afraid' of saying too little", (translated by the researcher).

better to do less but to do it right, i.e. respeak less information but make sure it is transcribed correctly, rather than produce lots of unintelligible content).

Overall satisfaction was high. For some of the students, the online setting interfered greatly with the output of the training, but they were all content with the workload and activities of the workshop. The satisfaction questionnaire also included a section on the experiment that will be referred to as the "Pre-experiment questionnaire" (Appendix 3), and will be presented in Section 3.4, Chapter 4).

The second questionnaire was submitted to participants after completion of the training and was aimed at monitoring their learning curves in the fields of MA, live subtitling and, more specifically, on the technique of respeaking. Given that all but one of the students had expressed little to no knowledge at all of MA or live subtitling and respeaking in the pre-workshop preliminary questionnaire (see Section 3.1.1), it was necessary to observe if and to what extent the training had had a positive impact on their understanding and awareness.

They were asked if they felt they had a better understanding of MA for auditory disabilities after undertaking the workshop. On a Likert scale ranging from 1 to 5 (again, 1 'strongly disagree', 2 'disagree', 3 'neither agree nor disagree', 4 'agree', 5 'strongly agree'), 14 students selected 4 or 5. The remaining student selected 3, which may still be considered medium-high in the range grades. They all explained in more detail that they felt a growing awareness as well as sensitivity towards the need for live subtitling, and that they were struck by the double accessibility that interlingual respeaking provides – both linguistic and for DHOHs.

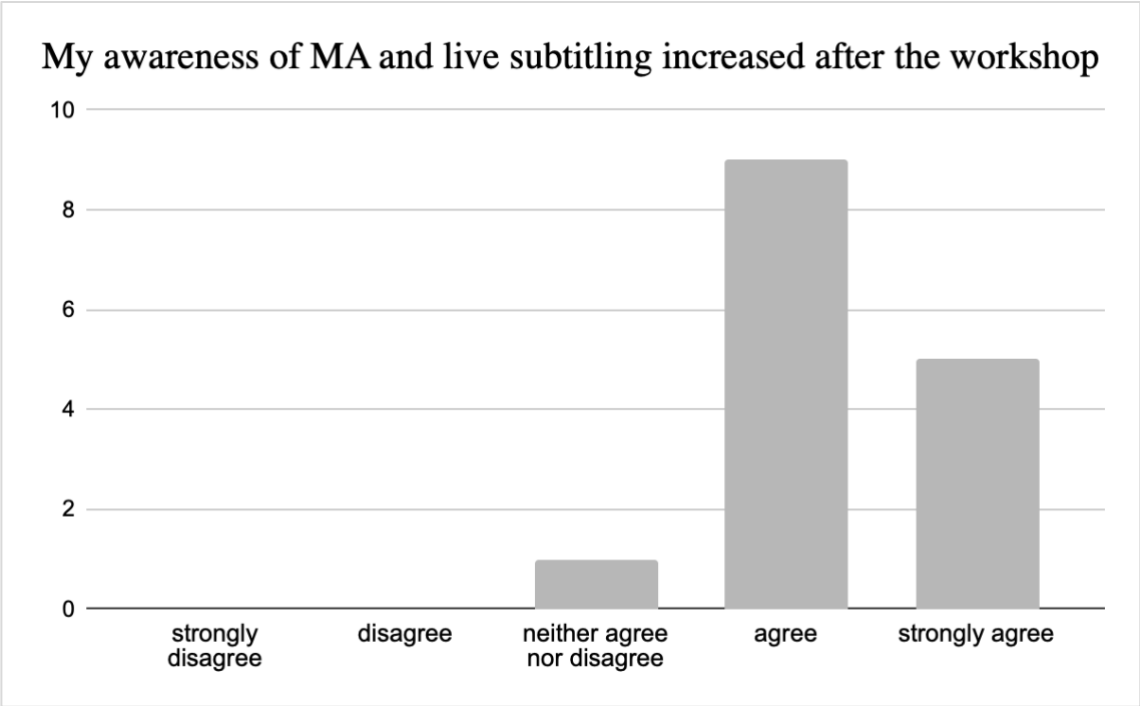


Figure 20 – Students' knowledge in the field of MA and live subtitling after the workshop

Nevertheless, it must be stressed that some participants observed that when practicing few occasions of specific features for SDH were included. The practice videos throughout the training, indeed, involved only one speaker, for example, therefore no speaker identification tags or labels needed to be introduced. Only some sound features such as background noises, sounds, or applause were included, while more exercise on that should have been foreseen.

4. Spanish to Italian workshop

A second edition of the interlingual respeaking workshop was delivered the following academic year 2021/22. This time, considering the potential of carrying out the same experiment on different methods of ILS and with a different language pair, the workshop was delivered from Spanish to Italian, despite knowing that the results from the two experiments could only be partially compared.

The second workshop started in November 2021 and lasted approximately 70 hours, spread throughout a three-month period, until the end of January 2022 (ten weeks training plus the experiment). Compared to the English to Italian workshop, this edition was marginally shorter since it was condensed to 5 teaching hours per week instead of 4 hours. The experiment was conducted throughout January and the beginning of February. Again, the training was delivered online via the university's official platform Microsoft Teams. Students attended remote lessons for five hours a week on average, and attendance was compulsory. Synchronous lessons amounted to approximately 50 hours of training, while the rest were dedicated to practical exercises and individual training. Its structure strongly resembled the previous workshop but with some differences. As noted in Section 3 of Chapter 2 ("Training modules in interlingual respeaking workshop"), during the first English to Italian experiment dictation to the SR software and segmentation were both major issues while respeaking (*ibid.*). Therefore, in the Spanish to Italian workshop the use of SR software and dictation practice were introduced sooner than previously and as part of a dedicated module. It was hoped that this would better strengthen the students' competences and give them the chance to train their voice profiles more robustly (Dawson, 2020: 232).

The workshop comprised a theoretical introduction to MA together with pre-recorded subtitling that was developed asynchronously but discussed altogether in class. A review of SI was provided and preparatory exercises given before moving on to use of the Dragon SR software to allow students to create and input their voice profiles through dictation exercises. Then, intralingual and interlingual respeaking were gradually introduced with a

theoretical framework on processes and skills. All modules and materials for the workshop are further detailed in Section 4.3.

4.1 Participants

The students that took part in this second experiment were the ones who participated in the Spanish to Italian workshop. There were 5 participants in total, 4 females and 1 male aged between 21-23 years old. All of them were students enrolled in the master's degree in Translation and Interpreting at the University of Genoa 3 of the students were in their first years and the remaining 2 in their second years Almost all of them had some previous experience, albeit introductory, of subtitling, but few had previous experience in SI since they had only just begun their first year of the master's program. They were all native Italian speakers and knew Spanish as a working language: four of them also studied Russian as a C language, while one studied French. They had all previously obtained bachelor's degrees in either 'Language Mediation' or 'Modern Languages and Cultures', and they all had worked with Spanish. Despite not being certified through CEFR, all the participants had good levels of Spanish albeit to varying extents and with some differences in linguistic performance.

4.2 Pre-workshop questionnaire and activity

As per previous, generic background information on the participants was compiled through Wooclap and a preliminary pre-workshop questionnaire submitted via Google Forms.

In the first, more informal Wooclap activity, students were confronted with the same two pictures: one simultaneous interpreter and one respeaker (*ibid.*, Section 3.1.1). They were asked to identify which task the subjects were carrying out. Prompted this way, many of

them recognized SI in the first figure, and many recognized respeaking in the second. Some of them responded with generic 'subtitling' when referring to the second figure, still showing that they did notice a difference between the two tasks pictured. When asked to brainstorm some features of respeaking, even without properly knowing the technique, responses included 'simultaneous interpreting', 'subtitles', 'accessibility', 'transcription', 'dictated subtitles', and 'shift of an oral text into a written one', thus showing that the group had had more of an opportunity to approach the topic prior to training when compared to the previous year's group.

The Google Forms questionnaire submitted to collect personal information on the students is available at Appendix 6. In addition to questions concerning gender, age, bachelor's studies and working languages, it sought to investigate students' knowledge and levels of preparation on pre-recorded subtitling and/or SDH, SI practice, MA, and STTI and respeaking.

4.2.2 Participants' background knowledge in subtitling and SI

Concerning subtitling, four out of the five had previously attended an intralingual SDH course or an interlingual pre-recorded subtitling module during their bachelor's or master's courses.

With SI, 2 students had previous experience in SI training either in Spanish or one of their other working languages, and 3 had no experience in SI particularly but some experience in other types of interpretation (dialogue or consecutive interpreting) and had carried out propaedeutic exercises only.

4.2.3 Participants' background knowledge in MA, STTI and respeaking

Participants were asked to score their perceived knowledge of MA, before training began, on a Likert scale from 1 to 5 (1 'strongly disagree', 2 'disagree', 3 'neither agree nor disagree', 4 'agree', 5 'strongly agree').

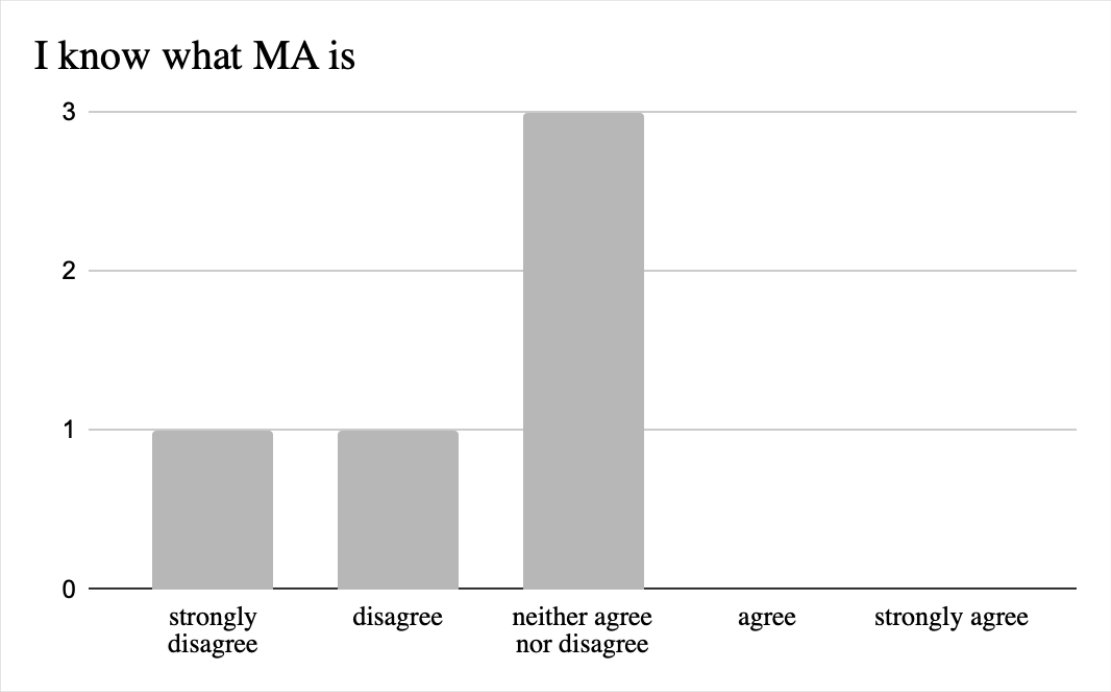


Figure 21 – Students' knowledge in the field of MA before the Spanish to Italian workshop

None of the students knew what MA was, but 3 responded that they had some idea but were not sure.

Concerning respeaking, as shown in Figure 22 below, 4 students had never heard of respeaking before, and therefore had no idea of what it was. Only one participant had heard of it but did not know exactly what it was.

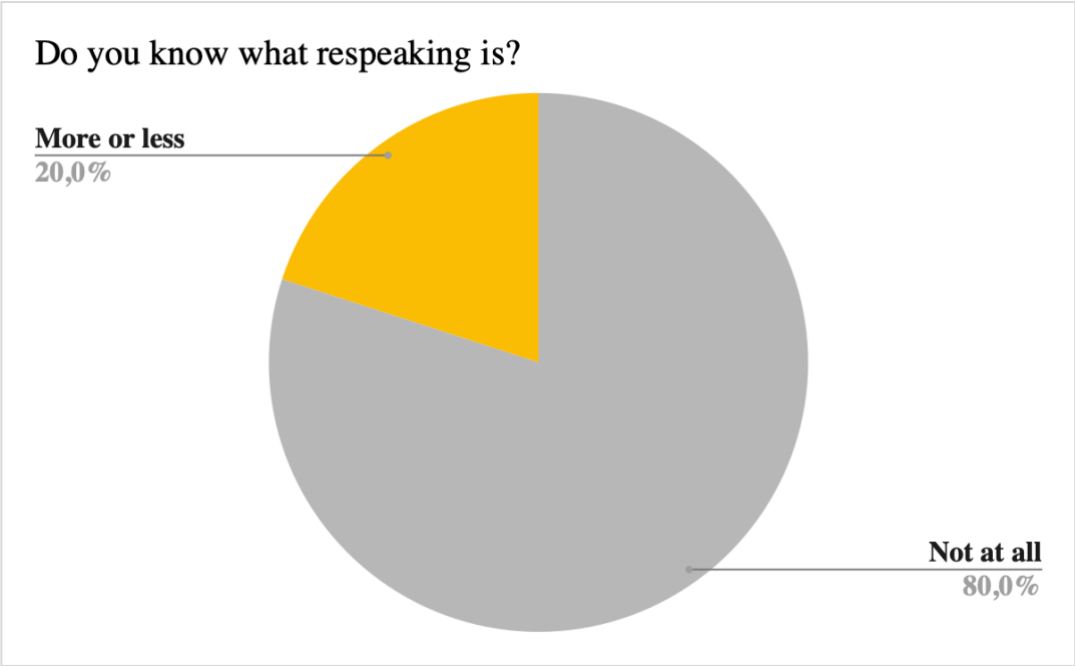


Figure 22 – Students' knowledge of respeaking before the Spanish to Italian workshop

4.3 Teaching methods and training modules

Thanks to the experience of the previous year, this second edition of the workshop was slightly modified according to the specific needs indicated by former students. Changes were implemented mainly via condensing the first module containing the introduction to MA including subtitling, and the anticipation of the module on the use of the ASR software to better cater with dictation strategies such as sentence segmentation and punctuation that were identified as major issues during the first experiment. The Dragon software processed transcription depending on students' dictation mode: as they were more familiar with interpreting rather than respeaking, they tended not to make any frequent pauses and segment their output to let the SR software process the information. The result was that some very long sentences were processed together and with many seconds of delay; this made them difficult to read and understand.

The materials from the ILSA course (video lessons and tutorials, reading, quizzes, etc.) were all in English, although because not all of the 5 participants spoke English fluently – and

given the workshop was indeed offered from Spanish – theory sections and informative videos were covered through explanation and discussion in class. As shown in Figure 22, a theoretical introduction to MA and live events accessibility was provided first, this time condensed with an introduction to subtitling. This module was covered asynchronously by students and later discussed together in class. Synchronous lessons started with Module 2, which focused on reviewing and strengthening SI, and Module 3 on ASR. Introducing use of the software beforehand was hoped to aid students' proper dictation when using it. Module 4 covered intralingual respeaking and, lastly, the context, background and practices of interlingual respeaking were covered. The training path finished with the experiment.

Another important change to the second edition of the workshop was the introduction of the NTR quality assessment model (Romero-Fresco & Pöchhacker, 2017) as a self-evaluation tool for the students when practicing. As highlighted by Dawson (2020), the use of assessment models during training in respeaking can help trainees become aware of their strengths and weaknesses. It is, indeed, an important moment for debriefing and checking one's performance, to fully take stock of the task with the goal of improving one's own performance. Under this premise, the students from the second experiment group were asked to carry out two NTR analyses of their respoken output. These were later commented on altogether with the tutor and classmates. Debating which grading to assign to different errors helped students become more aware of the impact that misrecognitions or mistranslations can have on readers.

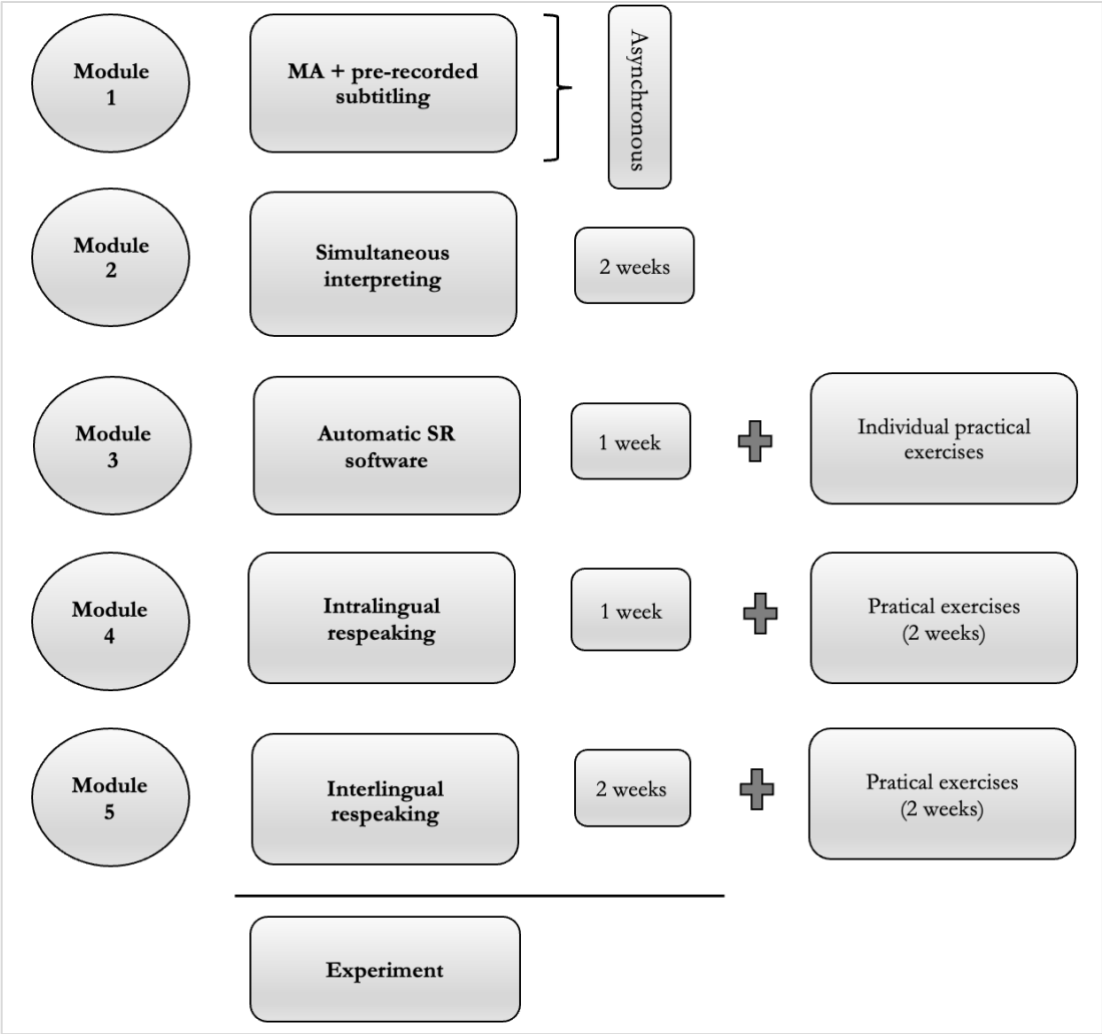


Figure 22 – An overview of the English to Italian interlingual respeaking workshop

4.3.1 MA and pre-recorded subtitling

This first module was covered asynchronously and autonomously by the students. As mentioned, this was a condensed module and included a section on accessibility services in AVT and an introduction to pre-recorded subtitling (see previous Section 3.2.2 "Introduction to pre-recorded subtitling" for further details) for the English to Italian workshop.

Module 1 – MA + pre-recorded subtitling
Unit 1: Human diversity and accessibility
1.1 Human diversity
1.3 Media accessibility and access services
1.4 Media accessibility and Audiovisual translation
Unit 2: Introduction to pre-recorded subtitling
2.1 Introduction to subtitling
2.2 Fundamentals of pre-recorded subtitling
2.3 Text condensation and line breaks in subtitling
2.4 Subtitling guidelines and shot changes
2.5 Subtitling exercise: learning to spot
2.6 Subtitling for the DHOH
2.7 Subtitling for the DHOH (II)
2.8 From pre-recorded to live subtitles: some hints and methods

Table 10 – Structure of Module 1 "MA + pre-recorded subtitling" of the workshop

4.3.3 Simultaneous interpreting

The structure of Module 2 was the same as was proposed for the first workshop (Section 3.2.3, "Simultaneous interpreting"). Preparatory exercises for SI consisted of counting, shadowing, paraphrasing and summarizing, and sight translation. The materials were this time proposed in Spanish and dealt with general topics without specific terminology. The Speech Repository and TED Talks platform were mainly used for training in SI.

4.3.4 ASR software

As previously highlighted, in this workshop use of the ASR software was introduced via an independent module – separate from intralingual respeaking practice – and was introduced sooner than when compared to the previous workshop. Particular emphasis was placed on dictation practice: trainees worked on tone and pronunciation, sentence segmentation, punctuation mark dictation, and error edition. They were also guided through training of the SR system with new words and word lists. The preliminary exercises using Dragon consisted of shadowing exercises with enunciation of punctuation marks, to get used to what would have been required of them while respeaking.

MODULE 3 – ASR software
Unit 1: Live subtitling and respeaking
1.1 Live subtitling and respeaking
1.2 Respeaking in Italy
Unit 2: Speech recognition and dictation
2.1 Speech recognition and dictation
2.2 Speech recognition engines + online tools
2.3 Misrecognitions
2.4 General dictation: punctuation
2.5 General dictation: word stress
Unit 3: Dragon Naturally Speaking
3.1 Dragon Professional Individual – Administrator Guide
3.2 Dragon Instructions and tutorials
3.3 Dragon – User guide
3.4 Customizing the SR software
3.5 Making live error corrections
3.6 Preliminary exercises with SR software

Table 11 – Structure of Module 3 "ASR software" of the workshop

4.3.5 Intralingual respeaking

The processes and practices of intralingual respeaking were introduced in Module 4. Students started working with Dragon, first on some ad-lib speeches by the tutor and then with some self-paced videos. At the beginning the speech rate given was low and gradually increased throughout. The same pre-recorded videos used as training material for this module were used in the previous workshop; please refer to Section 3.2.4 and Subsection 3.2.4.1 ("Intralingual respeaking" > "Training materials") for detail on video practice.

MODULE 4 – Intralingual respeaking
Unit 1: Intralingual respeaking
1.1 Initial intralingual respeaking
1.2 Split attention and rhythm in respeaking
1.3 Exercises in intralingual respeaking

Table 12 – Structure of Module 4 "Intralingual respeaking" of the workshop

4.3.6 Interlingual respeaking

Module 5 was the final module of the workshop and tackled interlingual respeaking. While the first two Units retraced the same content as the previous edition (refer to Section 3.2.5, "Interlingual respeaking"), Unit 3 also introduced the quality assessment calculation. Students applied the NTR model to calculate the linguistic accuracy of their respoken outputs, something that allowed for greater awareness of the impact of errors on audience comprehension. Preliminarily, some sight translation exercises from Spanish to Italian were proposed with enunciation of punctuation out loud. Then, students were tasked with exercises and practice in interlingual respeaking.

MODULE 5 – Interlingual respeaking
Unit 3: Interlingual respeaking
3.1 Sight translation (with punctuation) exercises
3.2 Exercises in interlingual respeaking
3.3 Quality assessment with NTR model

Table 13 – Structure of Module 5 "Interlingual respeaking" of the workshop

4.3.6.1 Training material

The rationale behind progressive exercises, from beginner/basic to intermediate level, was consistent with the previously proposed training. Pre-prepared speeches were first delivered by the tutor at a slow speech rate. At the beginning, the tutor recorded or delivered the speech at a much slower pace than the originals. This was the case for pre-recorded speeches by King Felipe VI, actor Antonio Banderas, former Uruguay President José Mujica, and novelist Isabel Allende. All speakers also had varying Spanish accents, which would have made it more complex for students to comprehend. After completion of the speech by the tutor, they also attempted to respeak the originals. The infamous commencement speech by Steve Jobs was also delivered but translated to Spanish despite being originally spoken in English. For longer videos, these were split and respoken in two parts, and for particularly fast speech rates (TEDx Talks "Lo positivo de fracasar en el amor" y "La pasión de ser diferente") students were allowed to watch once before respeaking, and then play them with a slightly reduced playback speed (0.75x), thereby lowering the speech rate to one more feasible for their level.

Title	Duration	Number of words	Words per minute (wpm)
Speech by King Felipe VI	00:13:05	960	74
Speech by Antonio Banderas, Premios Goyas	00:04:25	443	96
Speech by José Mujica, Río +20	00:07:40	936	128
Speech by Isabel Allende about 2006 Olympic Games	00:12:00	1,556	130
Speech by Steve Jobs (ES)	00:17:15	2,059	119
"Los peligros del teléfono móvil" Speech Repository	00:04:53	591	122
"El reciclaje de medicamentos" Speech Repository	00:05:40	748	140
"Lo positivo de fracasar en el amor", TEDx talk	00:08:03	1,338	167
"La pasión de ser diferente", TEDx talk	00:06:15	1,117	178

Table 14 – Training materials for interlingual respeaking practice

4.4 Post-workshop questionnaire

At the end of the workshop, a final anonymous questionnaire was given to participants via Google Forms (available at Appendix 7) to find out how satisfied they were with the different aspects of the training, and to monitor their progression in knowledge of the field of MA and respeaking after training.

As happened in questionnaires for EX1, also here, despite anonymous, answers could be partially conditioned by the students' desire of pleasing their tutor, avoiding commenting on downsides or negative aspects. The first 14 questions were dedicated to feedback on the overall training experience, on teaching methods, setting, and materials, as well as acquired

or improved skills. All five students declared they were very satisfied by the training in all aspects. When asked if they would continue training in interlingual respeaking given the choice, or if they would advise colleagues to undertake similar training, again they all marked 'yes'. Their responses are displayed in Figure 23 and are translated to English in the footnote below.

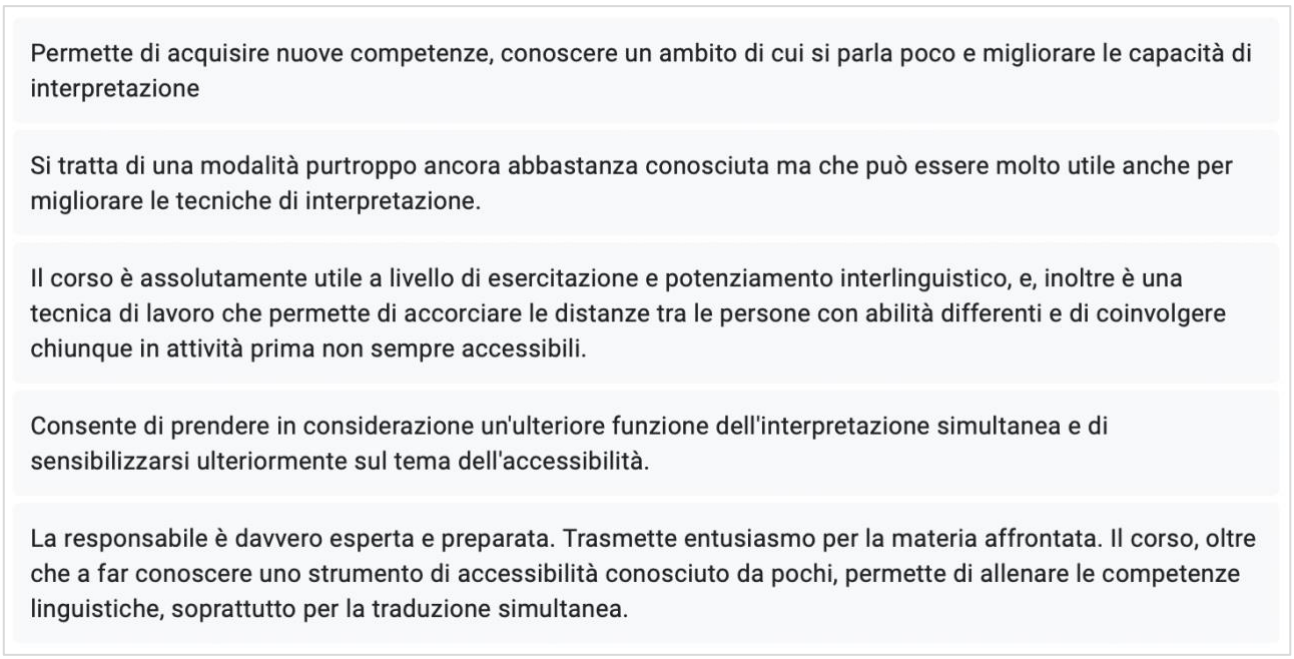


Figure 23 – Students' comments and opinions on the interlingual respeaking workshop[26]

[26] [The workshop] offers the chance of acquiring new skills, discovering a new field which is still relatively unknown, and improving one's SI abilities.

[Respeaking] is unfortunately still in its infancy and not widely acknowledged, but it can be very useful in improving one's SI abilities.

The workshop is undoubtedly useful concerning exercises and interlingual practice strengthening. In addition, it is a technique that allows for the reduction of barriers between people with different abilities and to involve anyone in activities that were not fully accessible before.

[The workshop] makes you aware of another function of SI and about accessibility services.

The trainer is professional and prepared, showing enthusiasm for the taught subject. The workshop not only educates you on an accessibility service that is still little-known, but also trains linguistic competences, especially for SI.

As in previous iterations, students highlighted the great similarity between interlingual respeaking and SI (more than subtitling), suggesting they could greatly benefit from parallel training, where both techniques are taught. More on this aspect of interlingual respeaking training in translation and interpreting master's degrees will be discussed in Chapter 5.

Concerning newly acquired competences or strengthened skills (in addition to the previously mentioned main competences of split attention, working under pressure, multitasking, problem solving, comprehension, linguistic competence, reformulation, etc.), this year's students noted: IT and specific software competences, [linguistic] improvement in L1 and L2, improvement of active listening comprehension, ability in grasping the most relevant information in a speech, better articulation and pronunciation. One student also mentioned an enhancement in their anxiety management in working in front of other people and with new techniques and tools. Interestingly, exactly as per the previous year's questionnaire, one student lauded the acquisition of the ability to deliver "a performance with as few false starts and hesitations as possible". They also acknowledged an improvement in their analysis ability through the NTR model, indicating that they increased self-awareness of their strengths and weaknesses in SI and respeaking.

The remaining 9 questions focused on monitoring the students' preparation in the field of MA through live subtitling and respeaking. Before training, half did not know what it entailed, and the rest were not completely sure (see Section 4.2.3). On completion, they all indicated that they were more prepared and knowledgeable of MA and live subtitling (between 4 and 5 on a Likert scale, as shown in Figure 24 below).

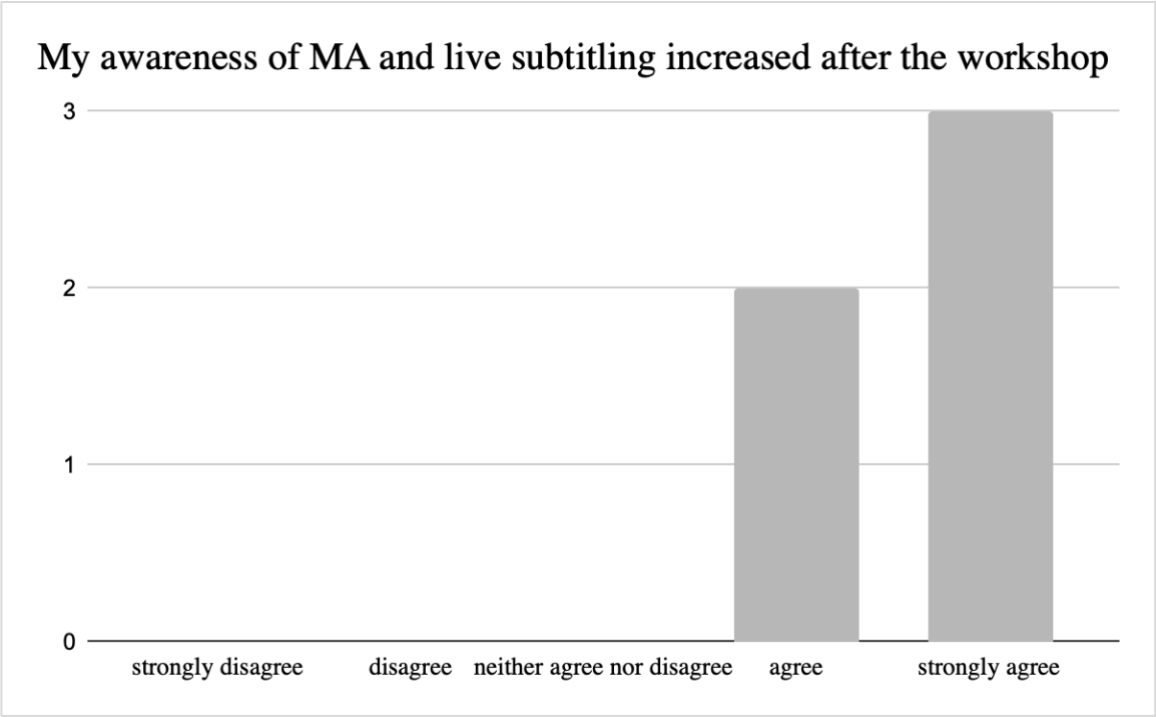


Figure 24 – Students' knowledge in the field of MA and LS after the workshop

In the second training as well the practice videos proposed involved only one speaker, and only some sound features such as background noises, sounds, or applause were included to be transposed as SDH features in the subtitles.

The post-workshop questionnaire also included a section on the experiment that will be referred to as the "Pre-experiment questionnaire" (Appendix 7), and will be presented in Section 4.3, Chapter 4).

Chapter 4.

Methodology and experiments setting

1. Introduction

After the necessary theoretical introduction of Chapter 1 and 2, and the recounting of the training experiences for the interlingual respeaking workshops in Genova, Chapter 4 finally introduces the relevant methodology for the two experiments of this research. The aim of this chapter is to provide the necessary information for the reader concerning the participants (who were the same as previously trained), the working teams, tested methods, and pieces of text (videos) used in EX1 and EX2, as well as the software and tools. After a short overview of similar research projects aimed at comparing different modes of producing ILS, this chapter is structured as per Chapter 3, by first outlining EX1 (English to Italian), and then EX2 (Spanish to Italian). For both, reference is made to the pre- and post-experiments questionnaires to track, among other things, their preparations and anxiety levels before testing, and any self-assessment or technical issues while performing.

2. Research on different ILS methods

Little to no prior research comparing quality of different ILS methods is available at the present. As previously mentioned in the research, this work was conceived and ran in parallel with other two projects: a case study from English to Spanish at the University of Vigo (Spain) by Pablo Romero-Fresco and Luis Alonso-Bacigalupe, and a Spanish to English case study at the University of Surrey (UK) by Hayley Dawson, both presented at

the IATIS (International Association for Translation and Intercultural Studies) conference in September 2021 together with the preliminary results of this doctoral research. In both experiments, all of the five methods tested in EX2 of this research were analyzed (interlingual respeaking; SI + intralingual respeaking; SI + ASR; intralingual respeaking + MT; ASR + MT), and their relevant findings and results will be discussed at the end of Chapter 5 making a comparison to this thesis results as well.

Pioneering research on quality of different ILS methods has been carried out by Carlo Eugeni (2020), focusing on HCI in ILS and investigating the role of technology in impacting ILS practices in terms of accuracy. To do so, Eugeni tested five different ILS forms during the 51st and 52nd Intersteno Congresses[27] focusing on time (or delay in the production of the transcribed text), accuracy and cost. The five different ILS modes were: 1) a simultaneous interpreter (Italian to English) plus a stenotypist; 2) an interlingual velotypist (German to English); 3) an intralingual velotypist (English to English) plus MT (English to French); 4) an intralingual respeaker (English to English plain subtitles) plus MT (English to French); 5) ASR (English to English) plus MT (English to French) plus live editor. Results concluded that ILS 1, 2 and 5 were not acceptable since they might involve too many professionals (Eugeni, 2020), or did not reach the required threshold by the IRA model. Among ILS 3 and 4, the latter involving an interlingual respeaker in the process was deemed to deliver the highest standard of quality despite a little more delay.

Another relevant study was conducted by Annalisa Sandrelli (2020), evaluating interlingual respeaking and SI as similar processes in a conference setting. The aim was to analyze the two techniques from English to Italian in terms of to what extent semantic content was conveyed in the interpreted feed and in the LS, showing as a result that greater semantic loss is encountered in subtitles, but that countersense and other meaning errors are balanced in both.

Another important research in the field was conducted by Claudio Fantinuoli and Bianca Prandi (2021), who evaluated the quality of a real-time speech translation (ST) engine

[27] For more information, see www.intersteno.org.

compared to the transcribed performance of simultaneous interpreters. Real-time or simultaneous ST, as defined by them, means a fully automated process in which the human disappears, namely Method 5 of EX2 of this research where the speech is recognized by ASR and then fed into a MT engine. Results from the experiment showed that human interpreters performed better in terms of intelligibility, while machine-driven processes performed slightly better in informativeness (semantic content) (Fantinuoli & Prandi, 2021: 245).

Reception of respoken and automatic subtitles has also been researched (Matamala *et al.*, 2015), even in comparison with the use of standard QWERTY method (Eichmeyer-Hell, 2021).

3. First main experiment (English to Italian)

The first experiment is also referred to as the 'main experiment' since it involved three times more participants than the second experiment – despite still working with small numbers – and because at the beginning it was the only one that was supposed to be carried out at all. Then, a second edition of the workshop with another language pair, and a growing interest in some other methods of producing ILS made it possible to run another smaller experiment.

The main experiment was conducted from English to Italian and was the culmination of the interlingual respeaking workshop that was offered in the previous months. As mentioned, the experiment was really one of the main objectives of the workshop itself: the aim was to introduce students to the practice of interlingual respeaking and familiarize them with it, to then participate in the experiment carrying out a respeaking task. It took place in February 2021 and was taken online, as per the training, due to the Covid-19 pandemic. The online setting posed many obstacles and technical problems for the organizing the testing, and more on this will be detailed in Section 3.3.

3.1 Participants

The fifteen participants that took part in this experiment were the students that were recently trained during the first English to Italian workshop in interlingual respeaking. When enrolling, all students were aware from the start that they would also participate in the doctoral experiment after training. All relevant information on the participants is available at Section 3.1 in Chapter 3.

3.2 Methodology

The experiment consisted of testing participants in delivering live subtitling through the three different methods that are detailed in the following section and then comparing them. In the experiment, they were asked to subtitle one video from English to Italian and, once the outputs for each method were collected, they were analyzed in order to answer the three reference RQs for EX1. In the assessment process for each output, two main variables were analyzed for a comprehensive quality assessment (Ofcom, 2013), namely:

- linguistic accuracy rate through the NTR Model (Romero-Fresco & Pöchhacker, 2017)
- delay in production of the subtitles

3.2.1 Tested methods

In addition to the other techniques previously mentioned to provide ILS in different contexts, interlingual respeaking has great potential to become one of the most interesting as it is gaining attention (Romero-Fresco & Alonso-Bacigalupe, 2022). According to scholars that have been working and investigating in this field for the last decade, possible suitable alternatives to interlingual respeaking (when concerning human-driven tasks) could be a combination of the different techniques with which it shares common ground (*ibid.*),

together with automatic systems. The methods tested in the present research are deemed appealing alternatives for the multilingual and DHOH accessibility market, combining intralingual respeaking and SI, ASR systems and MT software.

In the first experiment, three methods were tested and compared to answer EX1 RQ1, of which three could deliver a higher quality of LS. The methods involved were:

Method 1): Interlingual respeaking

Method 2): SI and intralingual respeaking

Method 3): SI and ASR

When measuring the extent of HMI in each, Method 1 is the most human-oriented mode (Eugeni, 2020), while Method 3 the most machine-centered.

All students were initially trained in interlingual respeaking when attending the workshop, but during the experiment only some of them performed it, while others were given different roles, specifically simultaneous interpreting and intralingual respeaking.
In Table 15, the number of participants needed for each method is outlined together with the role they performed: one for Method 1, two for Method 2, while for Method 3 no participants were required since Participant B's outputs from their simultaneous interpretations used for Method 2 were taken and fed into the ASR systems.

	Tested methods	Roles
1)	Interlingual respeaking	Participant A – interlingual respeaker
2)	SI + Intralingual respeaking	Participant B – simultaneous interpreter
		Participant C – intralingual respeaker
3)	SI + ASR	NA

Table 15 – Participants needed for each tested method

In the three modes, two variables were calculated in the assessment phase: linguistic accuracy through the NTR model calculation, and delay in the production of subtitles, i.e., how many seconds after the ST was pronounced were the subtitles displayed onscreen.
For Methods 1 and 2, delay was calculated in terms of the ending of each spoken sentence in the ST and appearance of the same sentence in the TT, while for Method 3 delay calculation was only speculative for technical reasons that will be better outlined in the next subparagraphs. Still, it was calculated starting from the time taken by the interpreter in translating, then adding extra time that would be the time for the machine to process the input.

For the three methods, Dragon Naturally Speaking voice recognition software (Version 15) was used in Italian, and for Method 3 an additional SR software, AppTek, was also used. All participants had had previous training using the software, and they were all provided with a professional headset with integrated microphone, as advised by the programmers themselves for best dictation and recognition. No subtitling software was connected to Dragon, thus the live subtitles were simulated. Participants were asked to record their performance through FlashBack Express recorder: the tool allowed for screen recording (for participants A and C) and for voice recording (participants B). In addition, screencasts in .mp4 formats could be easily exported separating the computer feed (ST audio playing)

and the microphone feeds, thus allowing for an .mp3 recording of the respeakers' performances to be isolated and listened to during the analysis stage.

Participants were asked to record the video source playing on a browser and the DragonPad positioned below the browser window, simulating the effect of live subtitles. The FlashBack Express exports allowed for tracking of mouse and keyboard movements on screen – signaled by a yellow moving dot – and for monitoring the respeaker editions as shown in Figure 25.

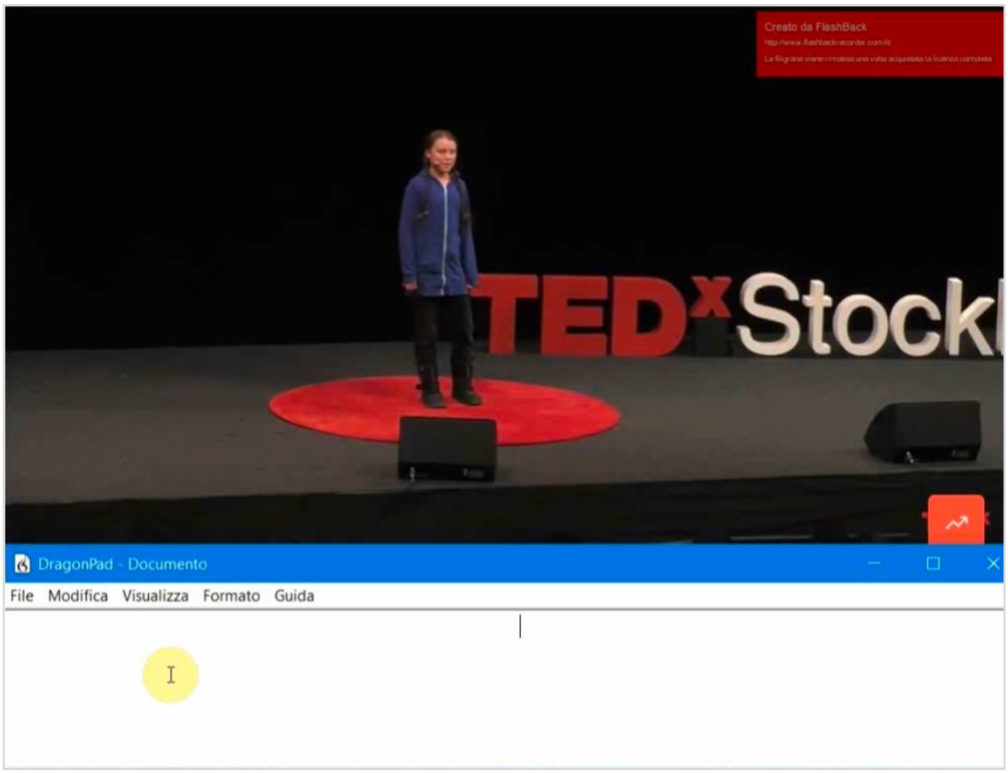


Figure 25. Participants' recorded screen example, with DragonPad below in EX1

3.2.1.1 Method 1: Interlingual respeaking

The first method is the most human-oriented, requiring only one professional (an interlingual respeaker) working alone. Exactly as for simultaneous interpreters typically working together in a booth in pairs, interlingual respeakers can also work together with a colleague, who usually edits the subtitles once created before they are sent on air. Working alone or with an editor largely depends on the working conditions (i.e., budget for example): if two people are working together, they can take turns with one carrying out the interlingual respeaking while the other checks and edits the subtitles created through the ASR software if necessary. Method 1 for this research was carried out with only one participant working alone, also because the task was under 15 minutes. Should the assignment exceed 30 minutes or so, then an optimal working condition would be of two professionals in pairs and taking turns as interlingual respeaking requires an increased cognitive load.
During the experiment. interlingual respeakers (Participants A) recorded their screens with FlashBack Express (all video and audio recordings for EX1 are available in Appendix 1). For Participants A, four different exported files are available: an .mp3 voice recording of the interlingual respeaker dictation and three different .mp4 videos (one with both computer and microphone feeds overlapping, one with only the computer/ST feed, and one with only the respeaker feed audible). All exports allowed for a thorough review of interpretation, dictation and edition during that phase of the analysis. After accuracy was calculated through the NTR model, delay was also quantified.

3.2.1.2 Method 2: SI and intralingual respeaking

The second method is even more human-oriented. It requires two professionals; one simultaneous interpreter (Participant B) who carries out the interlingual translation from English to Italian, and an intralingual respeaker (Participant C), who takes Participant B's output in Italian and respeaks it into Dragon to produce the subtitles. This method requires less of a cognitive load as the interlingual task is carried out by one professional, and the respeaking by another. It is important to note the possibility of increased costs and

therefore budget required for contracting two people instead of one as will be discussed more in depth in the discussion section at the end of the chapter.
Although an in-person setting for this method has never been tested before, the two tasks would be easily carried out in soundproof booths, the respeaker taking the interpreter's output as a kind of *relay*[28]. As the experiment was run online and participants were based in different locations, they connected via Microsoft Teams. The interpreter's feed was taken via Teams by the respeaker who, muted, worked intralingually with Dragon. Participants B and C had opportunity to rehearse the settings and modes of Method 2 during the last lessons of training, therefore this was not the first time they had attempted connecting via Teams and respeaking intralingually with a simultaneous interpreter's output.

During the experiment, interpreters (Participants B) recorded their performance and two different files were exported for analysis: an .mp3 voice recording of the interpreter's output and an .mp4 video of the ST video playing with the interpreter's feed, to make it easier to monitor *décalage* when having to calculate the delay from the ST playing and the subtitles to be created. Intralingual respeakers (Participants C) also recorded their screens during the experiment and, similarly to Participants A, four different file exports are available for their performances: an .mp3 voice recording of their dictation and the three different .mp4 videos as previously detailed. All exports allowed for a thorough review of interpretation, dictation and edition during the analysis phase. After accuracy was calculated through the NTR model, delay was also quantified.

[28] Defined by the EU Commission also as "indirect interpreting", *relay* is when interpreters work from a language they might not have in their combination through a bridging language (e.g.: interpreting from English to Spanish via a first interpretation to Italian) (European Commission, Conference interpreting – types and terminology: https://ec.europa.eu/info/departments/interpretation/conference-interpreting-types-and-terminology_en).
Here, independently from knowing the SL or not, the Italian feed is taken to dictate it intralingually to the software, completely relying on the EN>IT interpretation from Participant B, and without accessing the ST.

3.2.1.3 Method 3: SI and ASR

Method 3 is the more machine-oriented method, requiring fewer human-driven tasks and relying more on the technology. It requires one professional (a simultaneous interpreter) but for the experiment no participants were needed since Participant B's output from Method 2 was used in this case. Initially, the interpreter's translated Italian feed was to be live detected and transcribed by the ASR software with no human intervention and automatic punctuation addition also provided. To date, there are programs available that offer automatic punctuation, and some others allow automatic translation to several languages (Eugeni & Bernabé, 2019), but unfortunately this was not the case for Dragon version 15 which, at the date of the experiment, did not have automatic punctuation features available for Italian. This version of Dragon was only able to add punctuation marks automatically when transcribing recorded .mp3 files, not during live transcription. For this reason, for Method 3 the machine could not produce subtitles that were live recorded as per the first two methods. As an alternative, the interpreter's recorded feed was first exported in .mp3 and secondly fed into two ASR systems, thus allowing for only speculative delay calculation in the production of the subtitles, adding to the interpreter's *décalage* two extra seconds, which is on average how long a SR machine takes to produce the written text. This method allows for a lower cognitive load for the human since the voice recognition is processed by the machine and the output is not edited or checked.

3.2.1.3.1 ASR software: Dragon and AppTek

For Method 3 (SI + ASR), the participants' B outputs from Method 2 were fed into the SR engines. For a more informed and comprehensive result, two separate ASR systems were used to carry out the subtitling for analysis. The software used was Dragon Naturally Speaking (as per the other methods in order to be able to compare the outputs) and AppTek[29]. AppTek is an online tool that, like Dragon, allows for automatic transcription

[29] https://www.apptek.com.

from recordings with the addition of automatic punctuation. It also features live SR but the dedicated API was not used for this purpose to guarantee that the AppTek outputs would be comparable to the ones carried out feeding recordings to the engine (not live) using Dragon.

3.2.2 Assignment of roles and creation of teams

The three main roles (simultaneous interpreter, interlingual and intralingual respeaker) were assigned according to the participants' ability and performance in the different tasks during training. During the workshop, as seen in the previous Chapter, each participant was trained in and carried out SI, intra and interlingual respeaking practices and was evaluated on each. Those who usually performed better in SI were assigned the role of simultaneous interpreters, those who obtained a higher score in intralingual or in interlingual respeaking were consequently given intra or interlingual respeakers roles respectively. Students were also asked which role they felt more comfortable in performing according to their level.

Nevertheless, the tutor was aware that performing 'better' in one task did not automatically imply that a student would always perform to a higher level than the others in that role, or that all participants assigned to one specific role were equally proficient at it since this depended on the participants themselves. Another bias was any stress or anxiety prior to performing, especially during the final experiment. This is something that could have hindered and lowered the students' performances although this is something that cannot be easily measured or even avoided.

After the fifteen participants were assigned their roles, they were divided into five teams. As the objective was comparing the output quality obtained with each method under equal conditions, the teams were created with the most homogeneous performing level possible, despite the bias mentioned above. To do so, the best performers in SI, intralingual and interlingual respeaking during the training were assigned to Team 1, the slightly less well-

performing were assigned to Team 2, and weaker performers were accordingly assigned to the Teams 3, 4 and 5. The result was five teams of three participants to test the three different methods, as displayed in Table 16 below.

Teams	Participants
Team 1	Participant 1 (A)
	Participant 2 (B)
	Participant 3 (C)
Team 2	Participant 4 (A)
	Participant 5 (B)
	Participant 6 (C)
Team 3	Participant 7 (A)
	Participant 8 (B)
	Participant 9 (C)
Team 4	Participant 10 (A)
	Participant 11 (B)
	Participant 12 (C)
Team 5	Participant 13 (A)
	Participant 14 (B)
	Participant 15 (C)

Table 16 – Teams creation according to different roles

3.2.3 Experiment material

The material used during the experiment consisted of one video clip (referred to hereafter as 'text chunk') which all participants were asked to live subtitle according to their different roles. Participants A carried out an interlingual respeaking task, while participants B and C, in pairs, interpreted and performed an intralingual respeaking assignment. The

text chunk chosen was a full original speech in English with no specific terminology but with use of simple and familiar climate change-related lexicon addressing issues like global warming and perils threatening the planet, as could be easily inferred from the title itself. The speech was delivered by climate activist Greta Thunberg for a TEDx Talk in Stockholm in 2018. The mode of speech delivery was good, presented very clearly and with a standard speech rate. Greta Thunberg is not an English native speaker but little or no difference was perceived in her articulation, pronunciation or fluency during the speech. Concerning information density, some figures and percentages were presented, but the speech was not particularly dense. The length exceeded 11 minutes and it was one of the longest videos students had respoken; this could have slightly influenced their performance in terms of concentration fatigue, although this is unlikely and would have had very little impact on the final results. The video did not include any specific feature for SDH apart from some laughs by the audience and final applause. Note that, if respeakers did include it in the TT, no error was marked.

Title	Duration	Number of words	Words per minute (wpm)
"School strike for climate, save the world by changing the rules[30]"	00:11:03	1,369	125

Table 17 – Experiment material

Participants respoke (intra or interlingually) or interpreted the ST only once, and there was no opportunity to listen to the video beforehand, so at the time of the experiment it was the first time they ever heard of it. They were allowed to train the Dragon software with a

[30] The video of the live speech delivered can be retrieved at: https://www.ted.com/talks/greta_thunberg_school_strike_for_climate_save_the_world_by_changing_the_rules/transcript?language=it on the TEDx official website.

few words (acronyms, one term and a proper name) that were given to them in the SL, so that they could check if they were included in the software's vocabulary already, or if they needed to be added and trained. All participants created a voice profile on their personal software license that was used when training in class.

As shown in Table 18 below, each team provided 4 outputs (1 for interlingual respeaking, 1 for SI and intralingual respeaking, 2 for SI and ASR since two systems were used, Dragon and AppTek) for a total of 20 texts to analyze. The aim was to identify only one best scoring output for each team according to one of the methods, and the relative results of the experiments can be observed in Chapter 5.

	Teams of 3 people	**Methods**	**Outputs**
Video 1	Team 1	Methods 1, 2, and 3 (AppTek + Dragon)	4 outputs
	Team 2		4 outputs
	Team 3		4 outputs
	Team 4		4 outputs
	Team 5		4 outputs
			20 outputs

Table 18 – Overview of methodology for the first experiment

3.3 Online setting, organizational bias and solutions

Due to the Covid-19 pandemic, many aspects of the experiment were readjusted. Face to face learning would have allowed for a unique license purchase for the Dragon software product with multiple logins for the various students. This way, the serial numbers could have been installed in the university computers, requiring only the creation of new voice profiles by the students. However, Dragon Professional Individual licenses had to be

purchased so that they could be installed in each participants' computer terminal at home. This represented an issue as the students did not all have the same computers: some were compliant with the software requirements (memory RAM, processor cache, free space, etc.) but some were not, leading to slower information processing and transcription. This would have been the case, if only had there been any possibility of running the experiment in class using equally performing computers.

Headphones with integrated microphone were also purchased by the Department to provide participants with the same listening and, most importantly, dictation conditions.
Method 2 also brought technical issues. The interpreter and the intralingual respeaker had to virtually connect to each other so that the latter could take the former's output. This had to be done online via Microsoft Teams. While this was not a problem for the interpreters, but for the respeakers the workload of three different applications (Teams, Dragon and FlashBack to record their performance) running on the same machine sometimes caused performance issues. On these occasions, the Dragon recognition and processing (delay in the transcription) were affected most.
To minimize students' anxiety related to technical problems during the experiment, they could rehearse on two occasions before testing, to check Wi-Fi connection stability and performance of their computers with the three applications running simultaneously.

3.4 Pre-experiment questionnaire

The experiment for this research took place shortly after completion of training. Together with the post-workshop satisfaction and control questionnaire (detailed in Section 3.3, Chapter 3), a pre-experiment questionnaire was also given to participants, available at Appendix 3.

The first questions were dedicated to investigating the students' approach towards their participation in a research experiment for the doctoral thesis. Almost none of them had

previously taken part in research testing, but all expressed strong desire in seeing the experiment environment and believed it would have additional value to their educational experience. Some of the answers concerning their expectations of the experiment and the research are shown below in Figure 26.

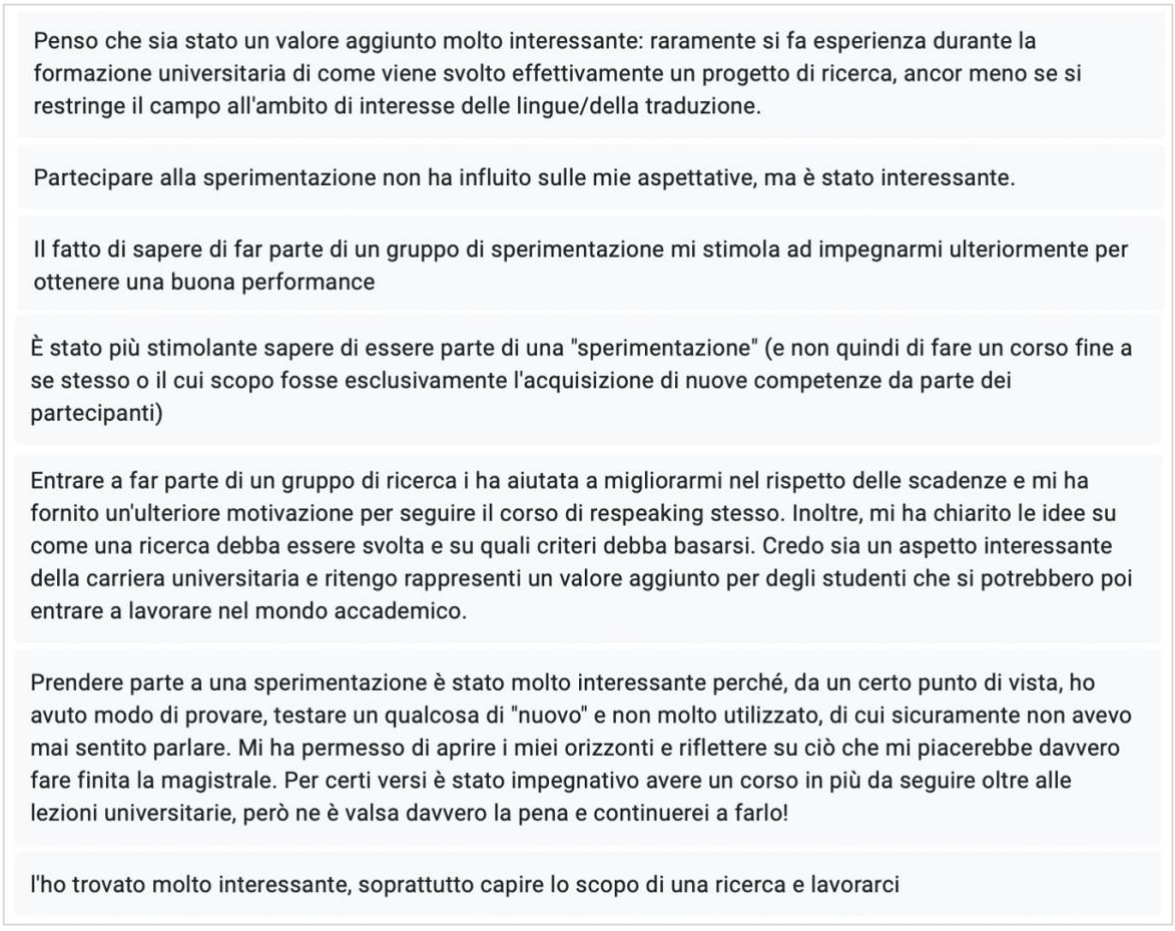


Figure 26 – Some impressions on taking part in the research experiment[31]

[31] I think it was very interesting and had added value: you rarely have the chance at university to experience how a research project is developed, even less in the fields of languages and translation.

Taking part in the experiment did not influence my expectations, but it was interesting.

Knowing that I was taking part in an experiment group has pushed me to work harder to perform well.

Interestingly, some felt motivated and stimulated by the forthcoming testing while others were not influenced by it at all. In their responses, it was underlined that this was the very first time they had been confronted with research methodology and that it was interesting and useful to see what it entailed.

When asked if they felt adequately prepared for the test, participants replied 'More likely yes than no', or 'Yes', demonstrating their confidence.

To observe their levels of anxiety about performance during the experiment (since they knew that the outcomes would be used to extract data), it seemed important to also investigate their feelings towards the test. When asked to briefly comment on how they felt when thinking about it, many expressed that they had some level of performance anxiety, agitation and/or fear. 8 out of 15 also declared that they were confident but that they would have to concentrate a lot because they were dedicated to it and were curious about how it would go.

3.5 Post-experiment questionnaire

At the end of the experiment, one final questionnaire was given to the participants (available at Appendix 4). This was to gather their impressions on the following aspects: level of difficulty (topic, terminology, length, and speech rate of the video), technical

---

It was stimulating to know that I took part in an experiment, and therefore not attending a course for its own sake, or whose aim was exclusively the acquisition of new competences by the participants.

Taking part in a research group improved my time management skills in respecting deadlines and it was motivating for the course attendance itself. In addition, it made clearer to me how research should be conducted and on which criteria it is based. I think it is an interesting aspect of a university degree, an added value for students who could work in the academic world.

Taking part in an experiment was very insightful because in some way I had the chance to try out something new and not widely used, that I had never heard of before. It broadened my horizons and allowed me to reflect on what I would really like to do after graduating. It has been challenging to follow an additional course as well as normal lessons, but it was definitely worth it, and I would continue training in this field.

I found it very interesting understanding the research scope and working on it.

problems encountered, self-evaluation of their overall performance. 14 out of the 15 students responded to the questionnaire. The questionnaire was not anonymous: this way it was possible to track each participant's performance according to their answers (i.e., verifying if lower performing students, for example, encountered some technical problems, or perceived the video as very difficult) and be able to find a correlation between results and reasons.
The questionnaire is made up of questions about a second text chunk that participants carried out on the same occasion. The data and results are not included in this research though, since only the first text was analyzed.

Participants were asked to rate how difficult they found the test on a Likert scale from 1 to 5. There were also open-ended questions in which they could express their views on any technical problems encountered (software, Wi-Fi connection, audio input etc.) and comment on their performance level.
11 participants deemed the experiment an intermediate level of difficulty, while 3 of them deemed it difficult.

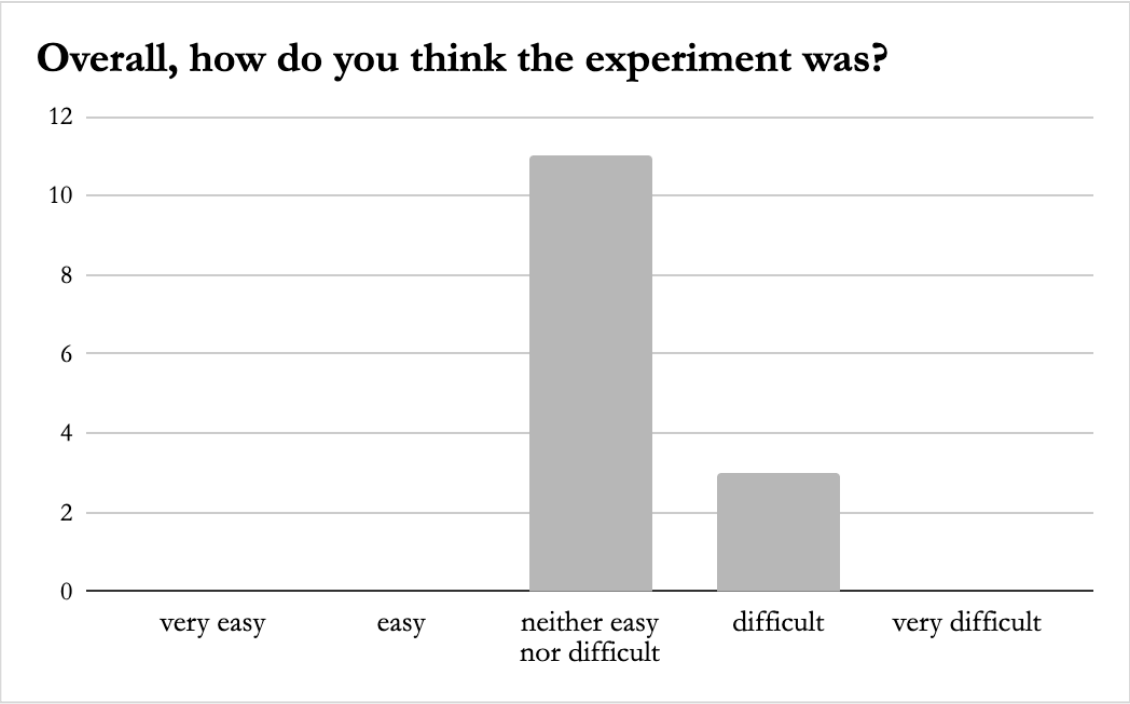


Figure 27 – Respondents' perceived level of difficulty during the experiment

On technical issues, some of the students declared they had had problems in listening properly, with the Wi-Fi connection stability, or PC performance. 9 students considered the speech rate to be of medium difficulty (neither too fast, nor too slow), and 1 considered it difficult. Overall, many pointed out that they condensed, reduced or omitted information here and there, and that they got confused when translating numbers and data. Some parts were identified as being dense in information, which gave them less capacity to monitor Dragon, for example. Despite the mode of delivery being deemed good by the tutor, one student pointed out that the intonation was very homogeneous and, thus, difficult to follow and understand, and the topic was generally recognized to be well-known but complicated. 7 students out of the 14 declared they were fully satisfied by their performance at the end of testing, or that they may have made some errors but were happy with the final result. Others said that they knew they had made many mistakes and that they could have performed better, so they were unsatisfied and confused, and that it was also difficult to follow the ST and consequently it was difficult to self-assess their own performances.

4. Second experiment (Spanish to Italian)

The second experiment of this research can only be partially considered comparable to the first. Firstly, the second was carried out using a different language pair (Spanish to Italian), and three shorter text chunks (instead of one longer chunk) were proposed; secondly, five different methods were tested and analyzed this time and, lastly, fewer participants were tested and received slightly different training that posed more emphasis on dictation strategies.

To summarize, the first three methods tested in the main experiment can and will be compared with the results extracted from the first three methods of the second experiment (see the final part of this thesis "Conclusions"), but bearing in mind the different language pairs and other variables. Otherwise, the two experiments also shared many similarities as well, such as training path, online setting, and mode of delivery.

In this case, the experiment itself was just one of the objectives of the Spanish to Italian interlingual respeaking workshop, so that trainees were familiarized with the techniques and could be tested on respeaking. The experiment was conducted from the end of January to the beginning of February 2022 as the conclusion of the workshop.

4.1 Participants

The five participants were the same students who had just completed the introductory training in interlingual respeaking. When the chance for the experiment arose, they were all invited to take part and accepted voluntarily. The participants' general background information is available at Section 4.1 in Chapter 3.

4.2 Methodology

The aim of the experiment was to test the participants' performance at different modes of live subtitling by comparing their obtained outputs. They were asked to perform three short videos from Spanish to English and, while carrying out the analyses for each method, the following variables were considered to answer the three reference RQs for EX2:

- linguistic accuracy rate through the NTR Model
- delay in the production of subtitles

4.2.1 Tested methods

In this experiment, five methods were tested and compared as per the other mentioned experiments (Dawson, 2021; Romero-Fresco & Alonso-Bacigalupe, 2021, 2022):

Method 1): Interlingual respeaking

Method 2): SI and intralingual respeaking

Method 3): SI and ASR

Method 4): Intralingual respeaking and MT

Method 5): ASR + MT

On a scale of different extents of HMI, Method 1 is the most human-centered mode, while Method 5 is more machine-centered and fully automated: the process is carried out entirely by the machine. As per the first experiment, all students were initially trained in interlingual respeaking whilst attending the workshop, but during the experiment only some of them

performed it while others were given different roles. The number of participants involved for each method is outlined in Table 19: one interlingual respeaker for Method 1, a simultaneous interpreter and an intralingual respeaker (Italian to Italian) for Method 2, and one intralingual respeaker but this time Spanish to Spanish for Method 4 whose output was then to be fed into the MT software. For methods 3 and 5, no participants were required since in the first case participant B's SI output from Method 2 was taken and fed into the ASR software, while in the second case the whole process was performed automatically by machines.

	Tested methods	**Roles**
1)	Interlingual respeaking	Participant A – interlingual respeaker
2)	SI + Intralingual respeaking IT>IT	Participant B – simultaneous interpreter
		Participant C – intralingual respeaker IT>IT
3)	SI + ASR	(same) Participant B – simultaneous interpreter
4)	Intralingual respeaking ES>ES + MT	Participant D – intralingual respeaker ES>ES
5)	ASR + MT	NA

Table 19 – Participants needed for each tested method

Dragon was used for each method, and during the experiment participants recorded their screens with FlashBack recorder, exporting all necessary files for the analysis. The participants' screen was set, once again, as shown in Figure 28 below.

Figure 28 – Participants' recorded screen example, with DragonPad below in EX1

4.2.1.1 Methods 1, 2, and 3

The first three methods were carried out following the same methodology as for the first experiment: an interlingual respeaker working alone, a simultaneous interpreter whose output was taken via Microsoft Teams by an intralingual respeaker for Method 2 and then fed into ASR software for Method 3. In this last case, the only technology used to produce the automatic transcription was Dragon, AppTek was not involved. For the first two methods, delay was calculated according to the Spanish Norma UNE 153010:2012, while for Methods 3, 4 and 5 it was only speculative. For further detail on Methods 1, 2, and 3 please refer to Sections 3.2.1.1, 3.2.1.2, 3.2.1.3.

4.2.1.2 Method 4: Intralingual respeaking and Machine Translation

This was the first time MT was introduced to the experiments. The system used was Google Translate, the neural machine translation service developed by Google. Similar experiments conducted in parallel with this research (Romero-Fresco & Alonso-Bacigalupe, 2021, 2022; Dawson, 2021) also used Google Translate, so it made sense to use the same system so that the results would be comparable to the others even with different language pairs. By introducing this method, Participants D were asked to carry out an intralingual respeaking task in Spanish despite none of them being native Spanish speakers. Concerning directionality (see Section 2.1, Chapter 3), it is not recommended that non-native speakers interpret or respeak to their L2 language, however this was the only way it could be done in this case as with native Italian speakers. As for the other participants, intralingual respeakers working in Spanish used Dragon, on which they had previously trained their voice profile in its Spanish version. The technology available did not allow for the ASR system to work together with the MT software and produce live transcription. For this reason, once the text chunks had been respoken, the DragonPad with the relative Spanish transcription was copied and pasted into Google Translate which had been set up for translation to Italian. The machine translated Italian transcription was then the final object of analysis for this method. In carrying out the process this way, delay in the production of the subtitles could only be speculative and not accurately measured.

4.2.1.3 Method 5: ASR and MT

The human component disappears completely in this last method and the whole process is carried out by machine. Dragon was used as the ASR system and Google Translate for MT. For the first part (ASR in Spanish) instead of switching on the Dragon microphone and playing the ST, letting it process and transcribe – which was tried out but provided barely usable outputs – the STs were recorded through FlashBack and fed into the software as .mp3 files. By doing so, the same conditions used for Method 3 were guaranteed, whereby the SI output was recorded and thereafter fed in. As per Method 4,

once the ASR process had been carried out, the text obtained was copied from the DragonPad and pasted into Google Translate which produced the output in Italian automatically. Delay could also only be calculated speculatively in this case.

4.2.2 Assignment of roles

As there were only five participants in this workshop and four people were needed to carry out the different methods as shown in Table 20, only one group was created to perform the different roles. During the workshop, each participant was trained in and carried out SI, intra and interlingual respeaking practices and was evaluated on all three. They also carried out an NTR analysis and self-evaluated their work for practice. Following the rationale of the first experiment, those more adept at SI were assigned the simultaneous interpretation roles, and those who performed better at intralingual or interlingual respeaking were consequently given a respective role. Roles were also assigned according to the participants' self-evaluation and self-confidence in performing them. Since there was one participant too many for the experiment group, two carried out the Spanish intralingual respeaking. Both outputs were analyzed, but only the one with the higher score was considered for the final results.

Roles	Participants
Participant A ES>IT	Participant 1
Participant B ES>IT	Participant 2
Participant C IT>IT	Participant 3
Participant D (I) ES>ES	Participant 4
Participant D (II) ES>ES	Participant 5

Table 20 – Team assignments

4.2.3 Experiment material

In this experiment, the participants were asked to attempt three short text chunks from Spanish to Italian. The texts were full original speeches in Spanish, two of which had no specialized terminology at all and featured very low levels of lexical and information density, whilst the other did feature some technicalities at syntactic level.

The first speech (deemed to be of low technicality and difficulty) had a slow speech rate of approximately 98 wpm and was given to the participants to warm up. It was a speech delivered by Pope Francis on the subject of the Earth's day and mentioned climate change and Covid. The Pope speaks a native Argentine Spanish that is dissimilar to the Castellano the students were used to. Nevertheless, the speech is delivered clearly and with good intonation and articulation. The second chunk was by ex-President of Spain Mariano Rajoy, addressing the Senate. It was faster (at 139 wpm), but much shorter in length and again did not feature any specific terminology. The third and final video was taken from the Speech Repository and was an old video by the European Economic and Social Committee. It was of medium length and at a standard speech rate, this time with a higher density of information and with institutional language (*dictamen, marco jurídico, directiva marco, propuesta de reglamento, autoridad, gestión del espectro, separación functional*).

The three videos did not include any specific feature for SDH with the only exception of the second one (Mariano Rajoy), for applause by the Members of Parliament. Note that, if respeakers did include it in the TT, no error was marked.

Title	Duration	Number of words	Words per minute (wpm)
Discurso del Papa Francisco para "El día de la Tierra[32]"	00:04:20	424	98
Discurso final como presidente de Mariano Rajoy a la Cámara[33]	00:01:24	208	139
Presentación del dictamen sobre el marco jurídico de las comunicaciones electrónicas[34]	00:02:35	318	126

Table 21 – Experiment material information

Participants attempted the task only once, and they were not given the opportunity to watch the proposed videos before testing, thus that was the first time they had seen them. They had the chance to train some words with Dragon, but only for the topic of Covid-19 and the surname 'Sánchez'.

[32] The video of the live speech delivered can be retrieved at: https://www.youtube.com/watch?v=LiTJvHmFtbE.

[33] The video of the full speech delivered can be retrieved at: https://www.youtube.com/watch?v=iSzk2Sm4Fl4.

[34] The video of the full speech delivered can be retrieved at: https://webgate.ec.europa.eu/sr/speech/marco-jur%C3%ADdico-de-las-comunicaciones-electrónicas.

As shown in Table 22, 18 outputs were provided: 6 for each text chunk (1 for the interlingual respeaking, 1 for SI and intralingual respeaking, 1 for SI and ASR, 2 for intralingual respeaking in Spanish and MT, 1 for ASR and MT). The relevant discussion concerning accuracy and delay for each will be detailed in Chapter 5.

Teams of 4 people	**Text chunks**	**Methods**	**Outputs**
Team 1	Text 1	Method 1	1 output
		Method 2	1 output
		Method 3	1 output
		Method 4	2 outputs
		Method 5	1 output
	Text 2	Method 1	1 output
		Method 2	1 output
		Method 3	1 output
		Method 4	2 outputs
		Method 5	1 output
	Text 3	Method 1	1 output
		Method 2	1 output
		Method 3	1 output
		Method 4	2 outputs
		Method 5	1 output
			18 outputs

Table 22 – Overview of methodology for the second experiment

Due to the enforced remote nature of the experiment the technological issues of PC performance encountered during the first experiment (Section 3.3) were also experienced here. Students had the opportunity of rehearsing using the different settings, especially for Method 2 where two people were connected remotely online.

4.3 Pre-experiment questionnaire

Some questions on the experiment were asked in the post-workshop satisfaction and control questionnaire (detailed at Section 4.4, Chapter 3) since it took place shortly after the conclusion of training. It cannot therefore be properly referred to as the 'pre-experiment questionnaire' since few questions were asked, but it aimed at investigating the following according to the students' own perspectives: levels of anxiety, expectations for the experiment, and preparedness.

When asked to comment on how they felt about the experiment, 4 expressed performance anxiety, agitation and/or fear, and only 2 declared they felt confident but that they would give maximum effort as they wanted to perform at their best. When asked if they felt adequately prepared for the test, 4 participants replied 'More likely yes than no', and 1 'Yes', demonstrating a certain amount of confidence.

4.4 Post-experiment questionnaire

The post-experiment questionnaire (available at Appendix 8) was given to participants shortly after the completion of the three text chunks. It aimed at investigating, for each text chunk, their perceptions of difficulty and the main issues and technical problems encountered, plus a self-evaluation of their overall performance. The questionnaire was not anonymous in order to allow for correlation of perceived difficulty with test output.

Questions on participation in the research experiment were asked in the pre-experiment questionnaire for the first experiment; here they were asked post-experiment. All of the participants perceived the experiment to be an adequate 'closure' of the training, helping them to monitor their improvement, and they found it useful. Their answers are shown in Figure 29 below.

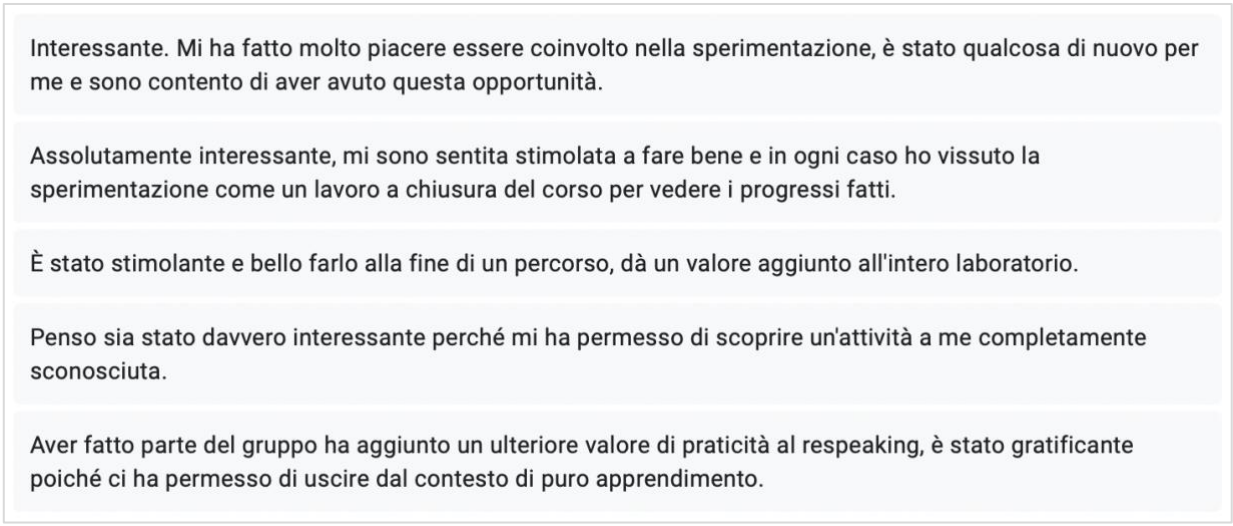


Figure 29 – Some impressions on taking part in the research experiment[35]

When asked to consider how difficult they found the entire test (all three videos), 2 participants chose 'neither easy, nor difficult', and 3 'difficult'.

[35] [It was] [i]nteresting. It allowed me to be more involved in the experiment and it was something new for me. I am glad I had this opportunity.

Particularly interesting: I felt stimulated to perform better and I view the final experiment as a closure of the workshop, to monitor our improvements.

It was stimulating and, really, to carry out [the experiment] at the end of the training path was of added value to the workshop.

I think it was very interesting since I discovered an activity that was completely unknown to me.

Taking part in the group has added an extra value to the practical side of respeaking. It was rewarding because it allowed us to go over the mere learning sphere.

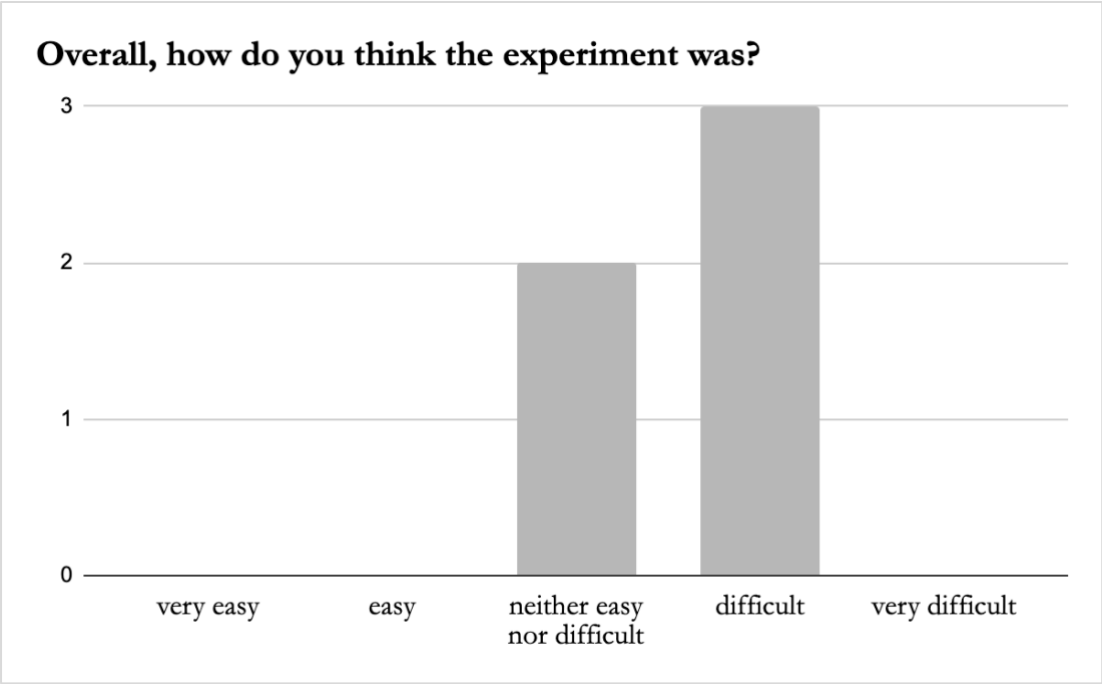


Figure 30 – Respondents' perceived difficulty of the experiment

They were also asked how difficult they found each specific text chunk. For the first chunk (speech by the Pope), all responded 'neither easy nor difficult' to 'very easy'; the second (Mariano Rajoy) was rated 'very difficult' by 1 person and for the third (Marco jurídico de las Comunicaciones electrónicas) 1 person rated it 'easy', 2 'neither easy nor difficult', and 2 'difficult'. For all of the videos, the participants deemed neither the topics nor the terminology to be complex.

Among the technical problems they encountered, these included slow recognition and processing of sentences by Dragon and the relatively unknown subject matter (especially for the third video).

Curiously, some found the slow speech rate of the first video to be an obstacle, as a more consistent speed of delivery would have helped to catch the syntactic connections between sentences. It was noted, however, that the rapidity of speech in the second and third videos was still a problem.

Concerning the self-evaluation of their performances, the majority felt that they knew they could have performed better and they felt unsatisfied, very unsatisfied or even confused by the end of the task. Few of them, and only in some videos, declared themselves fully satisfied by their performance.

Chapter 5.

Results and discussion

1. Introduction

In this chapter, the quantitative results obtained from EX1 and EX2 are presented and discussed. In the first section, the results from analyses of EX1 are outlined, followed in the second half by those of EX2, while finally a comparison of the first three methods pertaining to both experiments will be carried out.
In both cases, accuracy results are shown first, including the NTR scores for all methods and participants as well as averaged results with the relative 10-point reference scale. Attention is given to frequent typology errors in translation and recognition errors. Then, the delay calculation is displayed, followed by some final thoughts.

2. NTR analyses

For both experiments (EX1 and EX2) analyses were carried out using the NTR model (Romero-Fresco & Pöchhacker, 2017). In reference to what has been detailed in Sections 4.3.1 and 4.3.3 in Chapter 2, translation (T) and recognition (R) errors are graded with different severities: minor, major and critical (MinT/MinR, MajT/MajR, CritT, CritR). Effective editions (EE) are also calculated in the overall assessment, namely when the ST was successfully reformulated, without loss of content and impact on meaning.

Recognition errors are not divided into any other subcategories, but they are attributed a minor (-0.25), major (-0.5), or critical (-1) grading according to the impact they have on the TT depending on the type of mistake. As with translation errors, minor recognition errors are usually a lack of punctuation, the addition of small words by the software such as prepositions or articles due to audio interferences, missing gender/number concordances, and missing capital letters. Major recognition errors are those that transcribe words or units that are nonsensical, presenting a clear problem to the audience possibly leading to not being able to retrieve the ST meaning. Lastly, critical recognition errors occur when, through a misrecognition, a different sensical word is displayed even if it was not the intended ST one. In order to be deemed a critical error, the misrecognition needs to somehow fit in the context and introduce new information in this way. If a wrong word were transcribed in a context where it is evident that it is a mistake and makes no sense, that would account to a major error thus recognizable as a misrecognition by the reader.

The focus while analyzing and having to grade the different errors was mainly based on whether communication actually took place and to what extent it was achieved or not in the communicative event (Pöchhacker, 2004). This consisted of determining how important the different errors were, namely what relevance they may have for the target audience, although further investigation on priorities and perception of the target audience is most important. Assessment for the different methods focused on whether the content from the ST had been delivered successfully rather than on single typos, missing gender/number concordances or missing capital letters that did not impact meaning or readability in our opinion, and that would probably remain unseen to the audience itself while reading. Therefore, also when analyzing recognition errors, minor errors were only counted if two or more of them occurred in the same sentence, for example, but not if they were rare, isolated cases. Additionally, in focusing more on the achievement of communication, it is important to highlight that the assessment did not always refer to dependent vs. independent idea unit contraposition to grade errors, given the complexity of having to evaluate interpreted outputs (see Section 4.5, Chapter 2). While the norm would be that, for example, the omission of an independent idea unit would account to a major error (-0.5 points), in negotiating meaning the assessment sometimes overlooked the

mere lexical and syntactic level, focusing more on meaning. In other words, to exemplify: if the ST presented a whole idea through two independent units that was to some extent repetitive – despite standing alone – then a respoken output in which just one of them was rendered could have been considered as strategic reformulation and condensation, thus an all-round effective edition.

In Tables 23 and 24 below, two examples of the NTR templates are shown, the first taken from EX1, and the second from EX2 > source text 1, to display how linguistic accuracy was assessed.

Source text (transcribed audio)	**Interlingual respeaking-based subtitles**	**Errors**
When I was about eight years old, I first heard about something called climate change or global warming. Apparently, that was something humans had created by our way of living. I was told to turn off the lights to save energy and to recycle paper to save resources. I remember thinking that it was very strange that humans, who are an animal species among others, could be capable of	Quando avevo otto anni mi sono imbattuta per la prima volta in quello che viene chiamato " cambiamento climatico" (o riscaldamento globale). (1) (Apparentemente era qualcosa creato dagli uomini con il nostro stile di vita). Mi veniva detto di spegnere le luci per risparmiare energia (e di riciclare la carta per risparmiare le risorse). Mi ricordo che fosse strano pensare che gli uomini (2)	EE: no relevant information is lost by omitting "global warming". 1. MajT (cont-omiss) (0.5): a full independent unit is missing. EE: despite the omission, interpreting the condensation is strategic. 2. MinT (cont-omiss) (0.25): the parenthetical element is omitted. EE: no relevant information is lost by

changing the Earth's climate. Because if it were, and if it was really happening, we wouldn't be talking about anything else. As soon as you'd turn on the TV, everything would be about that. Headlines, radio, newspapers, you would never read or hear about anything else, as if there was a world war going on. But no one ever talked about it. If burning fossil fuels was so bad that it threatened our very existence, how could we just continue like before? Why were there no restrictions? Why wasn't it made illegal? To me, that did not add up. It was too unreal. So, when I was 11, I became ill. I fell into depression, I stopped talking, and I stopped eating. In two months, I lost about 10 kilos of weight. Later on, I was diagnosed with Asperger syndrome, OCD and selective mutism. That	(che sono una specie animale fra le altre) potessero essere in grado di cambiare il clima del pianeta. Perché in quel caso (se stesse davvero succedendo) non staremmo parlando di altro oggi. (3) (Se) Si accendesse una tv tutto parlerebbe di quello, dai media alle radio (e nei giornali) (4) (come se ci fosse una guerra mondiale in corso). Ma non si sente nulla di tutto questo. Nessuno ne parla. Se bruciare combustibili fossili è (fosse) così negativo da minacciare la nostra esistenza, (come potremmo continuare come se niente fosse?) come mai non ci sono restrizioni? Perché non è illegale? Per me (non aveva senso) era tutto così surreale. Quando avevo 11 anni mi sono ammalata, sono caduta in depressione, non ho più parlato né mangiato. In pochi mesi (due) ho perso 10 kg. Più tardi mi è stata diagnosticata la Sindrome di	omitting "and if it was really happening". 3. MinR (0.25): the omission of the hypothetical does not impact comprehension. EE: no relevant information is omitted by translating "headlines" with "media" and not listing 'newspapers' since other two similar examples have already been given. 4. MinT (cont-omiss) (0.25): some information is lost but it does not impact comprehension or lead to a loss of meaning. EE: no difference is perceived using the present instead of the past tense. EE: strategic reduction and reformulation since the following two sentences express the same idea. EE: condensing the two source text ideas into only one does not lead to a relevant omission. EE: reducing "two months" in "pochi mesi"

basically means I only speak when I think it's necessary – now is one of those moments. For those of us who are on the spectrum, almost everything is black or white.	Asperger, del (5) (il) disturbo ossessivo complessivo (6) (compulsivo) e (il) mutismo selettivo. Significa che parlo solo quando importante, e adesso lo è. In quel momento (7) (Per noi che abbiamo l'autismo) tutto era (8) (è) un bianco o nero.	does not omit any relevant information. 5. MinR (0.25): the article is misrecognized. 6. MinT (cont-subs) (0.25): thanks to the collocation for OCD, the misspelling of "complessivo" is recognizable 7. CritT (cont-subs) (1): new, misleading information is introduced here, omitting the input of the sentence. 8. MinT (cont-subs) (0.25): the past tense instead of the present remains due to the error before.

Accuracy rate	
MinT: (4 x 0.25 = 1) (cont-omiss) x 2 (cont-subs) x 2 MajT: (1 x 0.5 = 1) (cont-omiss) CritT: (1 x 1 = 1) (cont-subs) Total: 1 + 1 + 1 = 3	MinR: (2 x 0.25 = 0.5) MajR: 0 CritR: 0 Total: 0.5
NTR accuracy rate: (169 + 23) – 3 – 0.5	

---------- x 100 = **98.4%** (6/10) 192 EE: 8

Table 23 – Extract from an NTR analysis by Participant 13 (interlingual respeaker), from EX1

The template for the analysis and the calculation of the accuracy is presented in three columns: the left column contains the source text transcription, that is, the audio input that participants received through their headphones. The column in the middle reports the obtained written output, i.e. the subtitles created through the different methods, while the right column is dedicated to the comments concerning the assessment and the error grading. At the bottom, a section is dedicated to the recap of the number of errors depending on their severity, and to the calculation. A final section which is not shown here is dedicated to further comments and impressions for an overall assessment that takes into account also delay in broadcasting, and dwells more on other important aspects (glitches, critical errors and their impact, etc.).

Source text (transcribed audio)	**Respeaking-based subtitles**	**Errors**
Hermanos y hermanas, en esta conmemoración del día de la Tierra siempre es bueno recordar cosas que nos decimos mutuamente para que no caigan en el olvido. Desde hace tiempo estamos tomando más	Fratelli e sorelle, in questa commemorazione del giorno della terra (Terra) bisogna sempre ricordare cose che (1) (ci) diciamo l'un l'altro perché non vengano dimenticate. (da molto tempo stiamo prendendo	1.MinR (0.25): a particle is misrecognized, without any impact on the TT. EE: the sentence is consistently and strategically linked to the previous one with no impact on meaning.

conciencia de que la naturaleza merece ser protegida. Aunque sea por el hecho de que las interacciones humanas con la biodiversidad de Dios deben hacerse con el máximo cuidado y con respeto, cuidar la biodiversidad, cuidar la naturaleza. Y esto en esta pandemia lo hemos aprendido mucho más. También esta pandemia nos ha demostrado qué ocurre cuando el mundo se para, hace una pausa, aunque sea de unos pocos meses. Y el impacto que esto tiene en la naturaleza y el cambio climático de una manera tristemente positiva, hace daño.	coscienza di una cosa) Cioè che la natura deve essere protetta. Anche per il fatto che le interazioni tra la natura e l'umanità (2) (la biodiversità di Dio) devono essere fatte con (la massima attenzione e con) rispetto (occuparsi della biodiversità, della natura). Questo abbiamo appreso con la pandemia. Questa pandemia ha dimostrato anche cosa succede quando il mondo si ferma (3) (fa una pausa, anche se solo per pochi mesi). L'impatto che ciò ha sulla natura e il cambiamento climatico (4) (in maniera tristemente positiva), fa dei danni (5) (del male).	2. CritT (cont-subs) (1): "biodiversidad" was confused with "humanidad", introducing new information. EE: despite the omission of "máximo cuidado", the reduction in reformulation is strategic. EE: a part that was already introduced shortly before is missing, but it was a reiteration. 3. MinT (cont-omiss) (0.5): the reiteration of making a pause is omitted, with no impact on meaning or comprehension. 4. MinT (cont-omiss) (0.25): the hint on the way it is a sad thing is missing, but with no loss of content. 5. CritT (cont-subs) (1): translating "damages" instead of the idea of 'pain' is misleading since we are talking about climate change.
Accuracy rate		

MinT: 2 x 0.25= 0.5 (cont-omiss) x 2 MajT: 0 CritT: 2 x 1 = 2 (cont-subs) x 2 Total: 0.5 + 2 = 2.5	MinR: 1 x 0.25 = 0.25 MajR: 0 CritR: 0 Total: 0.25
NTR accuracy rate: (78 + 8) – 2.5 – 0.25 ---------- x 100 = **96.8%** (2/10) 86 EE: 3	

Table 24 – Extract from an NTR analysis by Participant 1 (interlingual respeaker), from EX2

Translation error typologies according to the NTR error grading model (see Section 4.3.1, Chapter 2) were identified: errors of content and errors of form, the first being omissions (cont-omiss), addition (cont-add) or substitution (cont-subs) errors, while the second correctness (form-corr) or style (form-style). Recognition errors were also identified and assigned a severity according to their impact on the TT. For the calculation, all recognition and translation errors were totaled and subtracted from the total number of words in the subtitles (N: words + punctuation marks), then divided by the same N and multiplied by 100. A dedicated space for comments is also provided in the template, which can include observations not only on the percentage obtained with the NTR calculation and error severity frequency, but also on other factors such as delay in the production of the subtitles.

For both experiments, since the corpus of analyses was limited, data were gathered and extracted on an Excel calculation sheet to detect frequency for some error typologies more than others, considering each method.

3. Results of the first experiment

In Table 25 below, the final percentage scores for each participant of each method after the completion of the NTR analysis are displayed. Together with the score, each output was attributed a grade on a 10-point scale.

	Methods	**NTR score %**
Team 1	Method 1	**97.2%** (3/10)
	Method 2	**97.5**% (3/10)
	Method 3 – AppTek	**96.5**% (2/10)
	Method 3 – Dragon	**95**% (0/10)
Team 2	Method 1	**97**% (2/10)
	Method 2	**96.8**% (2/10)
	Method 3 AppTek	**95.6**% (0/10)
	Method 3 – Dragon	**93.8**% (0/10)
Team 3	Method 1	[**96.3**% (1/10)]
	Method 2	**95.9**% (0/10)
	Method 3 – AppTek	[**92.5**% (0/10)]
	Method 3 – Dragon	[**86.4**% (0/10)]
Team 4	Method 1	**97.4**% (3/10)
	Method 2	[**94.9**% (0/10)]
	Method 3 – AppTek	**94.6**% (0/10)
	Method 3 – Dragon	**92.9**% (0/10)
Team 5	Method 1	**97**% (2/10)

	Method 2	**97.4**% (3/10)
	Method 3 – AppTek	**97**% (2/10)
	Method 3 – Dragon	**95.5**% (0/10)

Table 25 – NTR scores from EX1[36]

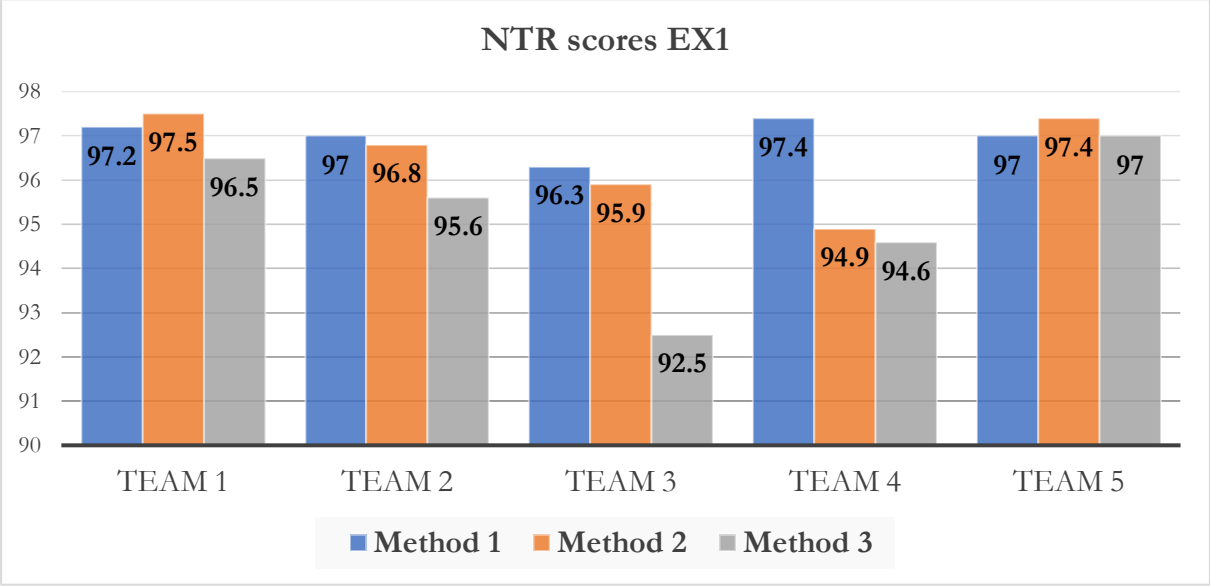


Figure 31 – NTR scores overview per method for the five teams (EX1)

Having worked with 5 different teams, each organized as detailed at Section 4.2.2 in Chapter 4, Table 26 also shows the best outputs obtained from the three methods by each team. We can see that Method 3 never particularly stood out, while Method 1 (interlingual respeaking) obtained the highest scores in Teams 2, 3 and 4; for Teams 1 and 5 this was Method 2.

[36] Percentages in squared brackets are not taken into account in the final average per method shown in Table 28 below, since those participants encountered technical problems during the test.

	Teams of 3 people	Methods	Outputs	NTR highest score
Text chunk 1	Team 1	Methods 1, 2, and 3 (AppTek + Dragon)	4 outputs	97.2% (Method 2)
	Team 2		4 outputs	97% (Method 1)
	Team 3		4 outputs	96.3% (Method 1)
	Team 4		4 outputs	97.4 % (Method 1)
	Team 5		4 outputs	97.4% (Method 2)
			20 outputs	**5 best outputs**

Table 26 – Best scoring output for each team

Table 27 below shows the average scores per method, distinguishing between the AppTek and Dragon engines in Method 3. Although all analyses were carried out regardless of technical problems, it is fair to mention that the .mp3 recordings by Participant 8 (simultaneous interpreter of Team 3) were very compromised. When fed into the ASR engines, Dragon especially struggled in processing the captions resulting in a very low score of 86.4%. Table 27 below features the average per method including this result, but it seemed fair to also show the results without including it due to the technical problems encountered (Table 28). Furthermore, Participant 12 (intralingual respeaker) of Team 4 also reported countless issues in listening properly to the interpreter's output in the headphones, resulting in a multitude of major omissions since the feed was inaudible throughout almost the whole text. In light of that, Table 28 features results excluding these two data, and the lowest ones by AppTek and Method 1 to guarantee the averages were calculated in the same conditions.

Methods	Average % per method		Rank
Method 1)	**97 %** (2.5/10)		1
Method 2)	**96.5%** (1/10)		2
Method 3) – AppTek	95.2%	**93.9%** (0/10)	3
Method 3) – Dragon	92.5%		

Table 27 – NTR scores average per method (I)

Methods	Average % per method		Rank
Method 1)	**97.2 %** (3/10)		1
Method 2)	**96.9%** (2/10)		2
Method 3) – AppTek	95.9%	**95.1%** (0/10)	3
Method 3) – Dragon	94.3%		

Table 28 – NTR scores average per method (II)

Method 3 largely depended on the way the interpreters enunciated translation. In Team 5, for example, AppTek obtained a 97% accuracy rate that, despite not reaching the minimum threshold, is one of the highest results. Listening to the recordings, Participant 14 of the

team, a simultaneous interpreter, had good enunciation and articulation of the interpretation leading to fewer hesitations and false starts.
Although the average is provided in the tables before between AppTek and Dragon results, for the sake of the research only the results by AppTek, which obtained higher scores, are taken into account since the software performed much better.

3.1 NTR analyses and linguistic accuracy results

After all outputs were collected, the NTR analysis (Romero-Fresco & Pöchhacker, 2017) for each of them was carried out to extract the results shown.
The NTR analysis for each participant of the teams was reviewed several times. The researcher completed the first round, which was validated by two other reviewers working separately who, despite not being professionals in the field, were both academic professors and interpreters. Instead of calculating the averaged difference between the researcher's and their assessment on specific errors, a compromise was reached between the two positions when they differed, even though few spots of different error gradings were detected throughout the whole analysis process.

All analyses templates for each participant are available at Appendix 1 of this research. A total of 20 outputs were analyzed in this phase, amounting to approximately 18,200 words. Since the total errors assessed for the 20 outputs was 2,019 errors, no statistical software was used to extract relevant data apart from Excel.
As could reasonably be expected, most of the translation errors were encountered in Methods 1 and 2, while many more recognition errors were encountered in Method 3 (both AppTek and Dragon).

Considering all methods – knowing that two analyses were carried out for Method 3 as two ASR machines were used for the same audio output – of all 2,019 errors 36.6% (739) were translation errors, and 63.4% (1,280) recognition errors.

Out of translation errors, 49.4% (365) were minor, 25.4% (188) major, and 25.2% (186) critical. Out of recognition errors, 70.8% (906) were minor, 23.7% (303) major, and 5.5% (71) critical.

In total throughout the methods and regardless of translation or recognition, 1,271 minor errors were detected, 491 major, and 257 critical errors. First, the translation errors for each method will be discussed and broken down into the different typologies and their frequency, followed by recognition errors.

3.1.1 Translation errors

A trend is confirmed when considering translation errors for all three methods, that is: regarding errors of content, omissions are predominantly minor errors, but also very frequent are major errors; additions are almost exclusively minor errors, while substitutions are almost entirely critical errors (see Figure 32 below). For Method 1, some additions were graded as critical errors since they introduced new information that was not originally in the ST. For the three modes, substitutions were largely critical errors with the exception of some minors where the content was not impacted by generalizations intended as substitutions, for example. In all cases, form errors in both correctness and style, despite accounting to very few occurrences, were exclusively minor errors. In the following sections and subsections, the different error typologies for each method are detailed.

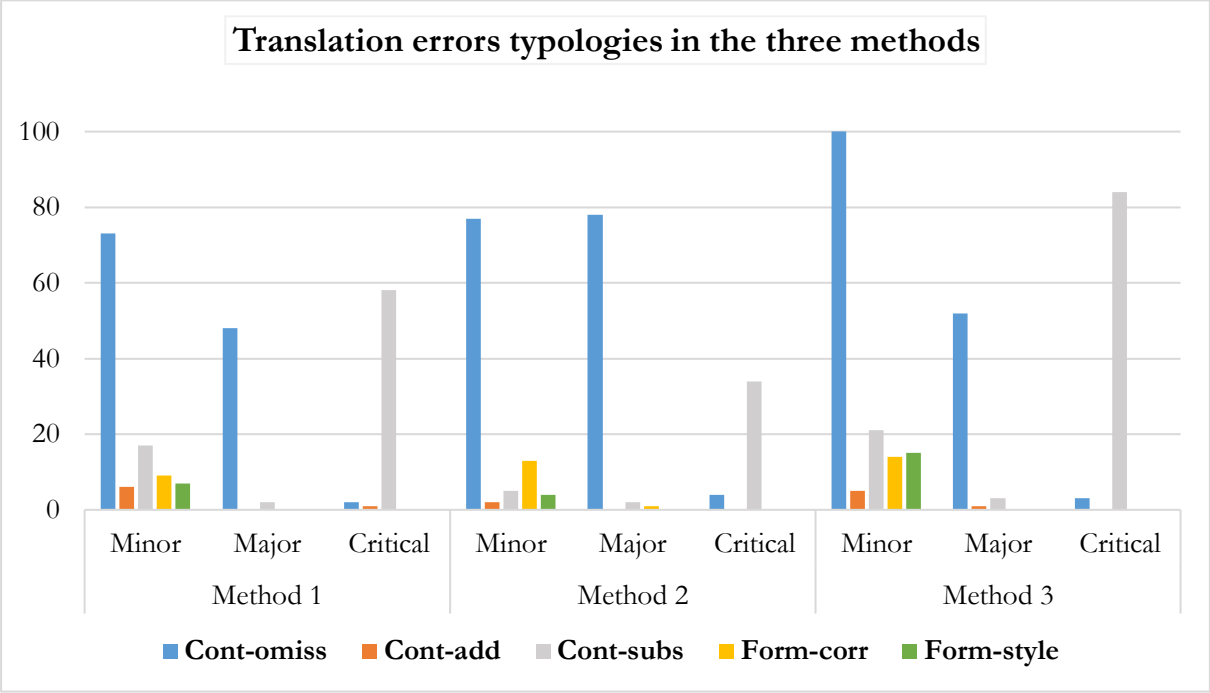


Figure 32 – Overview of the most frequent error typologies according to severity for the three methods

3.1.1.1 Method 1: Interlingual respeaking

Method 1 (interlingual respeaking) was the most demanding for participants, especially given they were students still being trained in SI, and not professionals. The results from the analyses of this method were still promising overall. Despite none of the participants reaching the threshold of 98% for accuracy, the performances that were delivered had some interesting solutions, although demonstrating that they still had to gain confidence with dictation and the SR software. Translation errors detected in Method 1 amounted to 30.2% of the total in all the methods (223 errors): of these, 50.2% (112) were minor errors, 22.4% (50) were major errors, and 27.4% (61) critical errors, showing a pattern that is repeated in Methods 1 and 3, where minor errors are much more frequent than major and critical (Method 2 was an exception, since minor and major errors were almost equal). Major errors should usually be more frequent than critical which should be

rarer but this was not the case as several translation problems were detected. Nevertheless, slightly too many critical errors were detected in each performance.

Among the 112 minor translation errors, 85.71% were errors of content (96), with a predominance of omissions (73, i.e. 65.18%), some substitutions and few additions, as shown in Figure 33 below. The very few minor errors of form correctness were mainly due to incorrect verb tenses (lack of subjunctive, for example), while minor errors of style were due to calques from English to Italian (e.g., "salvare energia" for "save energy", instead of "risparmiare").

Minor omissions were the most frequent errors and were mainly caused by the loss of a dependent idea unit (a complement specifying where, when, why or what) or were due to confusions in dictating verb tenses or prepositions. On some occasions, minor errors also accounted to short, repetitive units that were omitted in the live speech without losing content-driven parts.

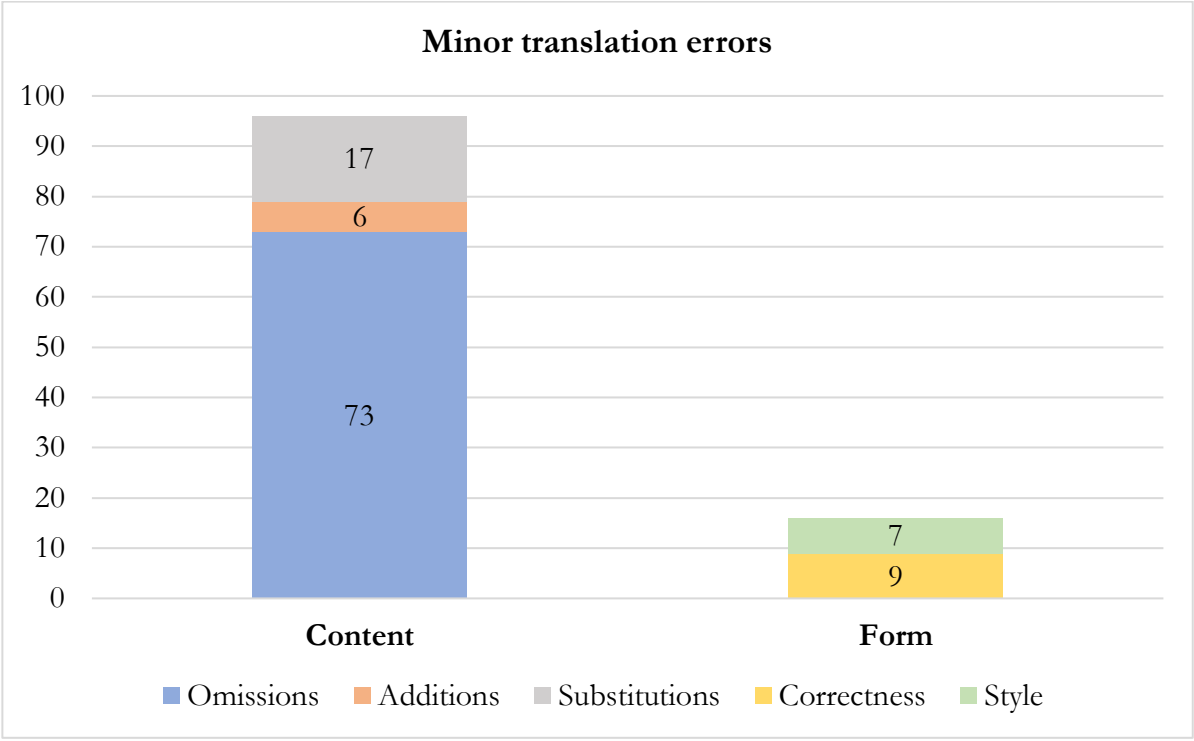


Figure 33 – Minor translation errors typologies for Method 1

The 50 major translation errors in Method 1 were all content errors, 96% of which were omissions (48). In the most human-oriented methods (1 and 2), major errors were almost exclusively due to omission of independent idea units, probably because participants did not hear or understand them properly. Naturally, no errors of form, either style or correctness, were assessed as major.

Critical translation errors for this method were 61 in total, with a predominance of substitutions (95.08% or 58 errors). Critical errors labeled as omissions were due to the loss of the second part of a sentence, which left the previous sensical one pending completion and introducing a completely different meaning altogether.

3.1.1.2 Method 2: SI and intralingual respeaking

Method 2 (SI and intralingual respeaking) was challenging in terms of technical setting, since two of the participants were connected remotely, one taking the other's output. The chance of encountering technical problems was high, and some students did find difficulties in managing Dragon recognition and Wi-Fi connection stability. In particular, Participant 12 (intralingual respeaker) struggled in receiving a clear audio input by Participant 11 (simultaneous interpreter) via Teams, resulting in many disrupted sentences and an unacceptably high number of independent units' omissions. The results of this team and method were still analyzed, but when averaging scores per method both figures with and without their scores were provided, since the same happened for another Method 3 test whose audio quality in recording was almost unintelligible (see final results Tables 27 and 28 above). In this case, neither performance achieved the 98% acceptability threshold posed by the NTR model, but the highest score obtained among all the others was indeed a Method 2, scoring 97.5% (followed by two 97.4%, one from Method 1, and the other from Method 2 again).

In the analysis for this method, since two participants contributed to the final output, errors (and EEs) by interpreters and respeakers were separated: errors made by the interpreter did

not feature any particular label since they were errors that would have been necessarily repeated by respeakers as well, who did not have access to the ST. Errors by the intralingual respeaker (Italian to Italian) can not be properly deemed as 'translation errors' since the task was an intralingual one and, therefore, an intralingual analysis through the NER model (see Section 4.2.2, Chapter 2) should have been carried out to assess this output. Still, errors by the respeaker that should be identified as 'Edition errors' as per the NER model are labeled here as translation errors "(RESP)", to isolate omissions, substitutions, or other form errors caused by the respeaker task only. Proper translation errors in the interlingual piece by the interpreter are labeled as "(INT)".

Translation errors detected in Method 2 totaled 218: 98 were minor errors (41 done by the interpreter and, consequently, by the respeaker as well, and 57 made only by the respeaker), 82 major errors (24 by the interpreter and 58 by respeakers alone), and 38 critical errors (22 by the interpreter and 16 by respeakers).

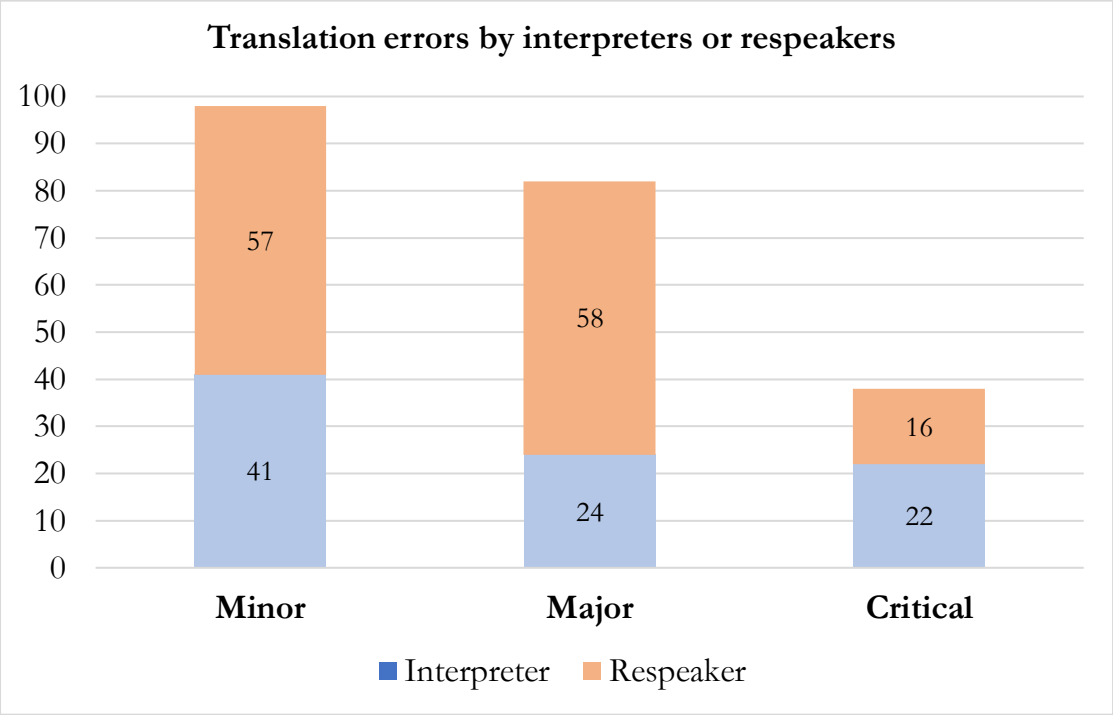


Figure 34 – Overview of translation errors by interpreters or respeakers in Method 2

As shown in Figure 34, among the 218 total translation errors for Method 2, 87 were made in the first place by interpreters, and 131 by respeakers only. The higher number of errors by respeakers was due to several factors, first and foremost to some technical problems with one team in particular in which the respeaker lost almost half of the interpreter's output, but also to the trend by intralingual respeakers to omit – either because they were lagging behind and they needed to keep up with the input, or because they failed in activating their short-term memory.

Concerning minor and major severities, the vast majority of errors were content omission errors. Among minor translation errors made by interpreters (41), 90.24% (i.e., 37 out of 41) were errors of content, with a high predominance of omissions (33, i.e. 89.19%), and among minor errors by respeakers 77.19% (i.e., 44 out of 57 errors) were again content omissions. Form errors by interpreters were on all occasions corrected by respeakers who, listening to a wrong concordance, or verb tense, automatically corrected them in Italian; still, 10 form errors among the 57 by respeakers were errors of form, probably due to their own confusion in dictating to the software (missing articles, wrong prepositions, etc.).

Among major translation errors by interpreters (24 errors) 95.83% (23) were content errors, of which 22 (91.67%) were omissions, and among the major errors by respeakers (58 errors) all of them were content based, of which 96.55% (i.e., 56 errors) were omissions.

Concerning critical translation errors, instead, the most frequent error typology was substitution. Among the critical errors by interpreters (22), 90.91% (i.e., 20 errors) were substitutions, and in errors by respeakers only (16), 87.50% (14 errors) were again substitutions.

Figure 35 below shows the main different categories according to the single error grading (minor, major or critical), regardless of who was responsible.

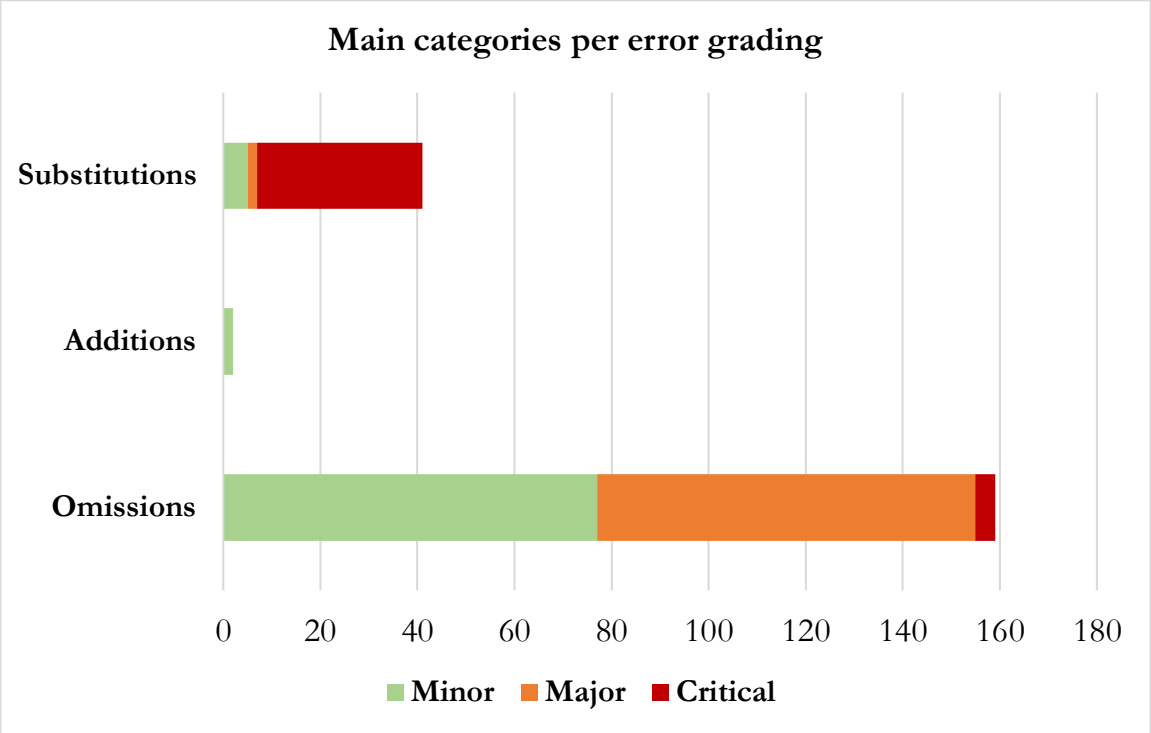


Figure 35 – Main categories per error grading for Method 2

3.1.1.3 Method 3: SI and ASR

Method 3 (SI plus ASR) ranked third after Method 1 and Method 2 respectively according to linguistic accuracy, both with AppTek that reached 95.9% accuracy, and Dragon with 94.3%. Nevertheless, Method 3 reached some high scores when compared to the other two methods, namely 96.5% and 97%, both with AppTek. As mentioned, the discussion will only refer to AppTek results since scored higher, but details on the error analysis are also given here in reference to Dragon, especially in reference to recognition errors discussed at Section 3.1.2 below.

Translation errors caused by the interpreter are then assessed in the final output of written text transcribed by the machines, therefore they contribute to the calculation of the accuracy. Since the same .mp3 recordings of the interpreters' outputs were fed in both machines, originally the same amount of translation errors should be present in both final

piece of texts. They are very similar indeed, even if in some cases some of them were omitted since the SR could not properly transcribe the audio.

In the whole method, 298 total translation errors were counted among the three severities: 155 were minor (79 in AppTek, 76 in Dragon), 56 major (27 in AppTek, 29 in Dragon), and 87 critical errors (46 in AppTek, 41 in Dragon). Some technical problems were also encountered with this modality, especially in feeding .mp3 recordings that had very low audio quality into the machines, resulting in incomprehensible outputs. In particular, data from Participant 8 (simultaneous interpreter) of Team 3 could not feasibly be taken into account for the final results since the SR software recognized almost nothing, or very few words in rare occasions. The results of this method and team were still analyzed, but when averaging scores per method both figures with and without the relevant scores were provided (see final results Tables 27 and 28 above).

Among the 155 minor translation errors for both machines, 81.29% were errors of content (126 errors), with a predominance of omissions (100, i.e. 64.52%). While additions were very few, 21 errors were represented by substitutions, and 29 errors (18.71%) were instead errors of form equally distributed between correctness and style (9.03% the first, 9.68% the second).

Concerning major translation errors, all of them were of content (56), 92.86% of which (i.e., 52 errors) were omissions, with very few additions and substitutions (only 1 and 3, respectively).

Critical translation errors (87 in total) were all of content, 84 of which (96.55%) were represented by substitutions.

As previously shown in Figure 35 for Method 2, the trend in Method 3 confirms the pattern in which additions are mainly minor errors, omissions are predominantly both minors and majors, while substitutions are almost entirely critical errors (see Figure 36 below).

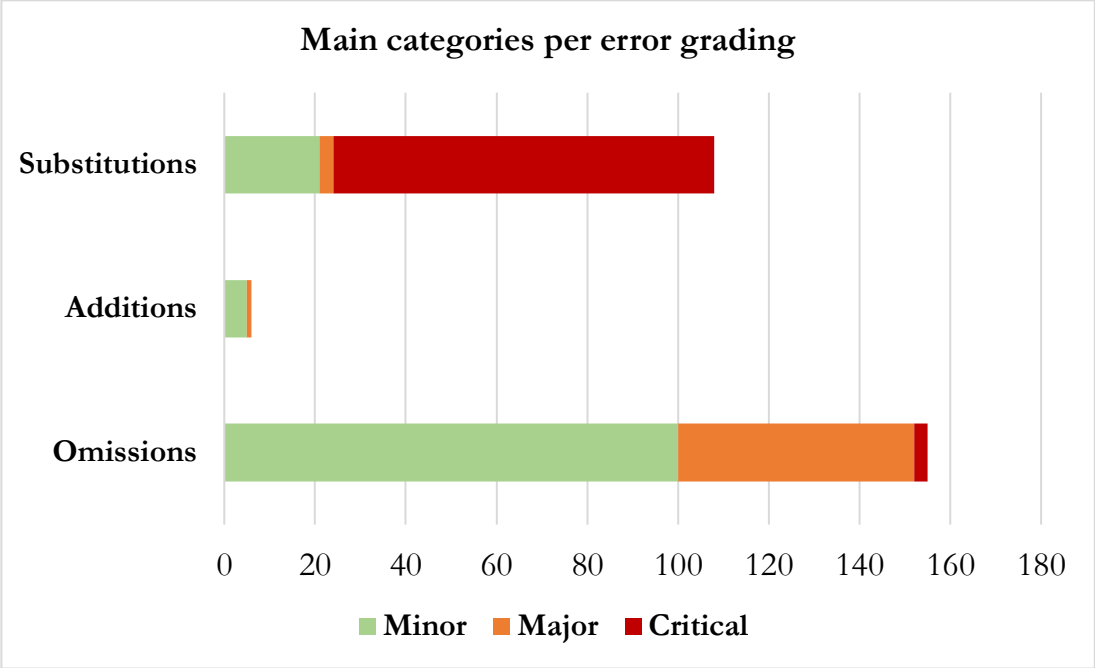


Figure 36 – Main categories per error grading for Method 2

3.1.1.4 Effective editions

Effective editions (EEs) are considered the parts of the ST that are strategically and successfully reformulated without loss of content and impact on meaning for the audience when displayed in the TT. As already mentioned, in the overall accuracy assessment, EEs are also considered even if it is impossible to determine if condensations, reformulations or strategic omissions are done voluntarily – thus they were a strategy adopted by the interpreter or the respeaker – or if they are due to a mistake. Either way, EEs were detected when for one reason or the other the text delivered coherent information, despite the coherency being reduced to some extent.

Table 28 shows all EEs for each participant and method of the experiment. Method 2 poses a differentiation between editions made by interpreters (INT) and, on some occasions taken over by respeakers as well, and those made by respeakers only (RESP) in the second part of the process, the intralingual respeaking task.

In a text chunk of 1,369 words in total, EEs were overall many and featured some interesting reduction and reformulation strategies by participants.

Method	Participant	Responsible/ Machine used	Total EEs
Method 1 – interlingual respeakers	Participant 1		21
	Participant 4		26
	Participant 7		27
	Participant 10		23
	Participant 13		30
Method 2 – SI + intralingual respeaking	Participant 3	INT	25
		RESP	4
	Participant 6	INT	19
		RESP	8
	Participant 9	INT	15
		RESP	6
	Participant 12	INT	10
		RESP	12
	Participant 15	INT	26
		RESP	6
Method 3 – SI + ASR	Participant 2	AppTek	26
		Dragon	26
	Participant 5	AppTek	20
		Dragon	19
	Participant 8	AppTek	11
		Dragon	3
	Participant 11	AppTek	19
		Dragon	18
	Participant 14	AppTek	28
		Dragon	28

Table 28 – Effective Editions per each participant of the three methods

EEs for Method 3 are those spotted during the interpretation process. Even for this method, in some cases not only the errors but also EEs could differ between AppTek and Dragon, since the two machines sometimes recognized different words, with Dragon especially failing in recognizing many idea units, therefore also losing some EEs.
Between Method 2 and Method 3, as the same simultaneous interpreter's output is taken for both, EEs detected in the transcription of the interpreter's output by ASR did not always correspond to those detected by the interpreters' feed recordings (INT) in Method 2, since on some occasions in this last case intralingual respeakers could have omitted some parts by the interpreters containing EEs.

3.1.2 Recognition errors

Recognition errors for methods 1 and 2 are referred to Dragon performance, since interlingual and intralingual respeakers worked with it. Errors for Method 3 are in reference to Dragon, and AppTek, since the interpreters' outputs were fed in both machines to compare their performance.

In Method 1, 72 recognition errors were detected in total: 59 were minor errors, 9 were major, and 4 were critical.
In Method 2, recognition errors totaled 44: 38 were minor, 5 major, and only 1 critical.
As shown in Figure 37, the third method was the one that registered the highest number of recognition errors. Data are shown separately: it is clear that Dragon failed on many more occasions in recognizing the audio input than AppTek.
As explained, the final scores concerning accuracy were displayed in Tables 27 and 28 both including Participant's 8 output who encountered several technical problems and excluding it from averaged calculation. Counting errors together with Method 3 AppTek and Dragon, 1,164 errors in total were assessed, divided as follows: 809 minor, 289 major, 66 critical errors. Without considering such output which was highly compromised by scarce audio quality, total errors detected were 883, 371 by AppTek and 512 by Dragon. This method

did not allow for human monitoring and edition and the simultaneous interpreter's output was directly taken by the machine, thus transcribing also all those oral features such as self-corrections, false starts and hesitations that are typical of SI.

AppTek (371 total errors) featured 285 minor errors, 63 major errors, and 23 critical errors, while Dragon (512 total errors) had 400 minor, 87 major, and 25 critical errors. Comparing the AppTek and Dragon performances, AppTek caused fewer errors and delivered a consistently better output of up to almost 2 percentage points than Dragon.

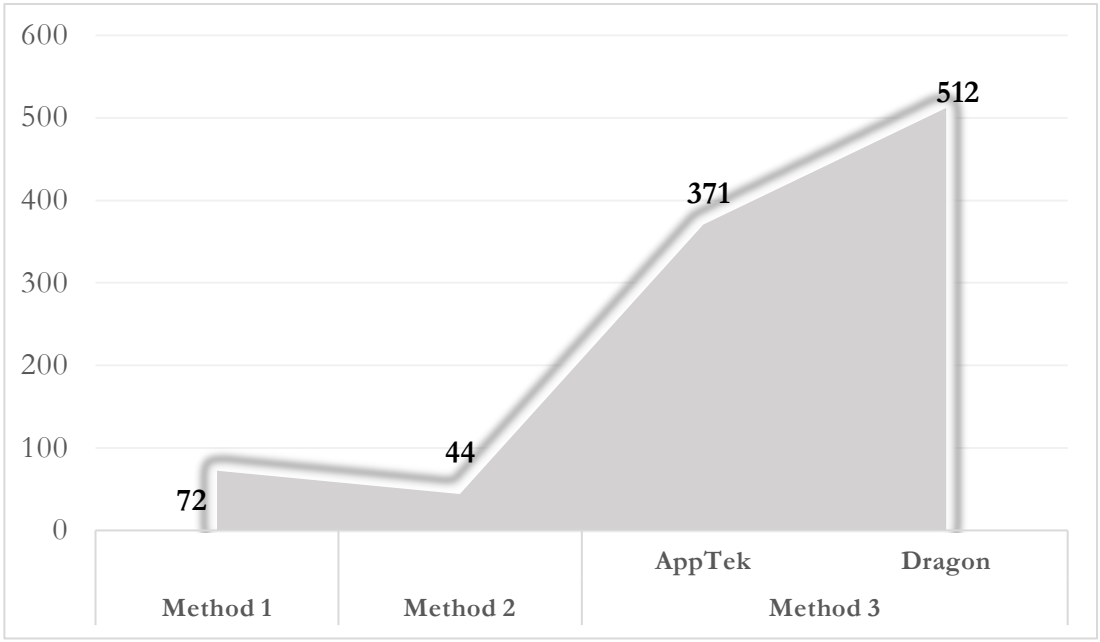


Figure 37 – Recognition errors in the three methods and comparison of AppTek and Dragon in Method 3

3.1.2.1Punctuation errors

All in all, the outputs fed into AppTek obtained quite higher scores than the ones processed by Dragon, which instead presented several problems. For both software, the main challenge was recognizing punctuation marks throughout the text, indeed several

errors were marked because of that, almost exclusively minors (see Figure 38 below). Furthermore, in AppTek a trend in the misrecognition of capital letters after full stops was also detected, although such errors were not taken into account for the assessment. As in the other two methods, if a capital letter was missing at the beginning of a new sentence, or in a proper name (e.g., 'terra' instead of 'Terra') it was not marked as a minor error.

Among the 285 minor recognition errors counted on four outputs (not five) for Apptek, 144 (i.e., 50.5%) were due to lack or exchange of punctuation marks (commas between subject and verb, lack of commas in lists, full stops instead of question marks at the end of questions or vice versa etc.). No major errors were due to punctuation errors, while interestingly among the 37 total critical errors, 3 were caused by punctuation.

For Dragon, among the 400 minor recognition errors 233 (58.3%) were errors of punctuation. Very few major errors were due to punctuation (3 out of 87, 3.5%), and among 25 total critical errors, 2 were punctuation-related.

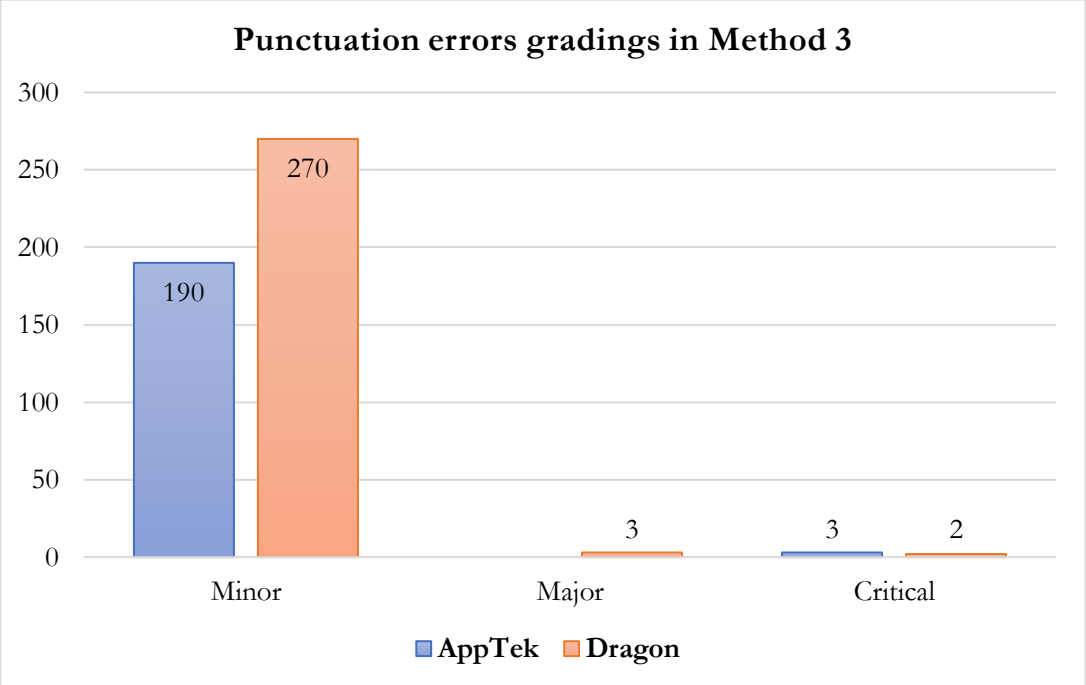


Figure 38 – Punctuation marks error gradings in Method 3, for both AppTek and Dragon

Any critical errors that were due to punctuation problems are particularly interesting, because they do not occur often. Following the idea that "commas can save lives" (ILSA, 2020), such errors were either because a full stop or a comma were placed where they did not belong, instead of a question mark, thus giving the sentence another meaning. When considering example sentences such as: "I find inspiration in cooking my family and my dog" instead of "I find inspiration in cooking, my family, and my dog", or "Let's eat grandma!" instead of "Let's eat, grandma!", noted below are the very few cases of misrecognition of punctuation that led to critical error gradings.

Example 1:

ST: "Are we evil? No, of course not. People keep doing what they do because […]".

TT: "Siamo cattivi, (?) ma non è questo, le persone fanno quello che fanno […]".

Back translation: *We are evil, but it is not it, people do what they do since […].*

Example 2:

ST: "Are we knowingly causing a mass extinction?".

TT: "Sappiamo di causare un'estinzione di massa. (?)".

Back translation: *We know we are causing a mass extinction.*

Example 3:

ST: "We all think we know, but we don't. Because how could we?".

TT: "Pensiamo di sapere. In realtà non sappiamo perché. Come potremmo?".

Back translation: *We think we know. Actually we do not know why. How could we?*

3.2 Delay

As anticipated in Chapter 4, Section 3.2.1 ("Tested methods"), a vital role in the overall quality assessment of the final outputs is also played by the delay with which they are broadcast. For the first experiment, delay was calculated with a random sample using the official Spanish guidelines on subtitling DHOH, the Norma UNE 153010 on subtitling for DHOH. It consists of choosing one sentence-ending at every minute of each source text to calculate the lag between the moment in which a specific oral utterance was spoken, and the moment in which it was displayed as subtitles on screen (Romero-Fresco & Alonso-Bacigalupe, 2022).

For Method 3, delay calculation was instead speculative since the intralingually respoken output was fed into the ASR system only at a second moment thereafter. For this, the delay (*décalage*) in translating by the simultaneous interpreter was calculated in terms of seconds, and then an average of 2 seconds were added speculatively as the time the ASR engine would take to process the input and produce the transcription. No differentiation was made between AppTek and Dragon since they would take an average of 1 second each for the processing, thus delay for Method 3 is included in only one result.

Table 29 below displays an overview of each participant's delay, while average per method is provided in the following Table 30.

Teams	Methods	Delay (seconds)
Team 1	Method 1	8.1
	Method 2	9.6
	Method 3	5.6
Team 2	Method 1	5.9
	Method 2	13.3
	Method 3	5.3
Team 3	Method 1	7.2

	Method 2	[18.7]
	Method 3	[7.7]
Team 4	Method 1	[34.9]
	Method 2	7.9
	Method 3	5.3
Team 5	Method 1	6.7
	Method 2	15
	Method 3	4.8

Table 29 – Participants' delay results in seconds

Methods	**Delay average (seconds)**	**Rank**
Method 1	12.6	2/3
Method 2	12.9	2/3
Method 3	5.7	1

Table 30 – Delay average per method (I)[37]

It is important to mention that in Team 4, Method 1, the interlingual respeaker (Participant 10) in the post-experiment questionnaire (and as it is clearly viewable in the reference screen recording), reported several technical problems while performing, first and foremost with the slowness of Dragon in processing captions, broadcasting them on the DragonPad in blocks of 3 sometimes 4 sentences altogether, thus leading to a compromised delay of 34.9 seconds. For this reason, it seemed fair to offer both averaged delays per method, with (Table 30) and without (Table 31) Participant 10's results: in order to offer the same

[37] Results in squared brackets are not taken into account in the delay average per method shown in Table 31 below, since those participants encountered technical problems during the test.

conditions to all the three methods for the second averaging, with only 4 instead of 5 teams, the longest delay results were cut off from the second counting for Method 2 and 3 also (that is, 18.7 seconds for Method 2, and 7.7 for Method 3).

Methods	Delay average (seconds)	Rank
Method 1	7	2
Method 2	11.5	3
Method 3	5.3	1

Table 31 – Delay average per method (II)

At a first glance, Method 1 never exceeded a 9' second delay, while Method 2 always scored around a 10-12 second delay, involving two participants working together, which naturally increases the time in which the task is carried out. Method 3 is the mode that ensured the least possible delay, ranking first among the three. In this case, in fact, apart from the *décalage* of the interpreter, delay is minimized by the very quick response of the ASR engine.

3.3 Final thoughts

The results from the first experiment from English to Italian, testing three methods (interlingual respeaking, SI + intralingual respeaking, SI + ASR), did not reach the minimum required threshold of 98%, and only some vaguely promising results reached 97 or 97.5% in some cases. There are many reasons as to why the accuracy threshold was not achieved, the main being that the participants were students and not professionals, and that they had only had an introduction to the practice of intra or interlingual respeaking, thus having to improve greatly in the technique before being able to perform properly.

Nevertheless, the ratio between the different tested methods is consistent since students were involved for all of them; they were not professionals in either respeaking or simultaneous interpreting. Technical problems played an important role as well, hindering the software performance in many cases and preventing some processes that could have run smoother if there were any possibility for a face-to-face setting for the experiment.

Overall, the results from this experiment corresponded to the ones extracted by the other two experiments carried out in parallel (Romero-Fresco & Alonso-Bacigalupe, 2022; Dawson, 2021). Method 1 provided the highest level of accuracy, and second highest concerning delay; Method 2 ranked second concerning accuracy, and third concerning delay, while Method 3 had the lowest level of accuracy, but featured a maximum potential synchronicity, i.e., minimum delay. From here, a relation between delay and accuracy can be detected: the shorter the delay, the less accurate the TT, which means in other words that outputs that scored higher in accuracy also took longer in terms of delay. This relates back to what was previously noted for an overall quality assessment: the shorter the delay, the better, but only provided good linguistic accuracy is delivered (*ibid.*). In searching for the best mode to use in real life contexts, it is useful to consider if a shorter delay is more important than the detriment of accurate subtitles.

Method 1, despite its intrinsic complexity, delivered higher scores in accuracy and allowed for human monitoring and edition of the output. Despite the fact that it could be even more challenging when used in specialized settings, it does seem a good compromise for now.

Method 2 also allows for good monitoring and edition since two professionals work together on the same task, although errors by the interpreters cannot be detected as such by intralingual respeakers, who are likely to repeat all of them in their task. Furthermore, delay is slightly longer and it is unlikely to become any shorter. As pointed out by Romero-Fresco & Alonso-Bacigalupe (2022), Method 2 can also be very interesting since it offers two different outputs depending on the audience's needs: audio for those who wish to listen to content, and subtitles for audiences who need (or prefer) written texts.

This is the same for Method 3, which is unfortunately not quite fully acceptable as a result of accuracy. ASR, still, is always evolving and will surely improve, most likely allowing for higher accuracy rates in a short time. Nevertheless, human monitoring and checking is not encompassed in this mode, which can be risky especially for those oral features of SI that the automatically transcribed text cannot avoid: false starts, hesitations, self-corrections, etc.

As a final thought, it seems that this is not merely a matter of finding the one best method amongst the tested three, but rather a matter of adequacy in respect of different settings. This also depends on the costs involved and other factors such as the level of technicalities and terminology involved, the reception of the audience, and language pairings.

4. Results of the second experiment

Table 32 below shows the results of Experiment 2, from Spanish to Italian. As seen in Chapter 4, in this experiment three shorter full speech videos were used to test the participants (ST1, ST2, ST3). The final NTR percentage scores for each of the five methods are displayed individually, and then on average. Together with the actual score, each output was attributed a grade on a 10-point scale.

Methods	ST1	ST2	ST3	NTR % average score
Method 1 Interlingual respeaking	96.1% (0/10)	97.3% (3.3/10)	97.9% (4.8/10)	**97.1%** (2.8/10)
Method 2 SI + intralingual resp.	99.1% (7.8/10)	98.7% (6.8/10)	98% (5/10)	**98.6%** (7.5/10)
Method 3 SI + ASR	97% (2.5/10)	98% (5/10)	96.6% (1.5/10)	**97.2%** (3/10)
Method 4 – Part. 4 Intralingual resp. + MT	96.7% (1.8/10)	95.1% (0/10)	95.2% (0/10)	**[95.7%]** (0/10)
Method 4 – Part. 5 Intralingual resp. + MT	95.3% (0/10)	97.7% (4.3/10)	98.5% (6/10)	**97.2%** (3/10)
Method 5 ASR + MT	94.4% (0/10)	96.7% (1.8/10)	95.1% (0/10)	**95.4%** (0/10)

Table 32 – NTR scores from EX2[38]

The following Table 33 shows the results cut down to NTR average scores, together with the rank obtained when comparing the five methods. Method 2 ranked first as it was the

[38] Results in squared brackets by Participant 4 are not taken into account in the delay average per method shown in Table 31 below, since those participants encountered technical problems during the test.

only one achieving over 98% and yielding a satisfying outcome, followed by methods 3 and 1 which were very similar with a difference of only 0.1 points. Method 4 came next, scoring lower by almost a full point, Method 5 ranked bottom with the lowest accuracy rate of below 96%.

Methods	NTR % score	10-point grading	Rank
Method 1 Interlingual respeaking	**97.1%**	(2.8/10)	3
Method 2 SI + intralingual resp.	**98.6%**	(7.5/10)	1
Method 3 SI + ASR	**97.2%**	(3/10)	2
Method 4 Intralingual resp. + MT	**97.2%**	(1/10)	2
Method 5 ASR + MT	**95.4%**	(0/10)	4

Table 33 – NTR scores ranking

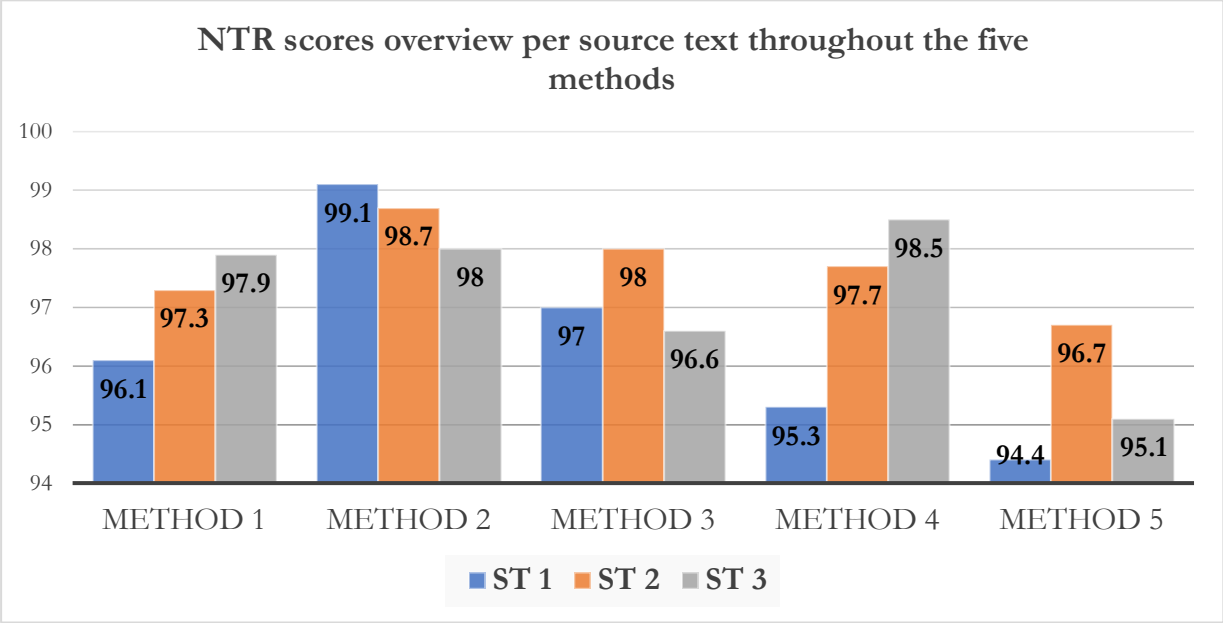


Figure 39 – NTR scores overview per method throughout the five teams (EX2)

4.1 NTR analyses and linguistic accuracy results

The analyses were carried out using the NTR model by the participants of the experiment themselves. They each carried out their own analyses for the three text chunks they attempted followed by a peer review. A final review was carried out by the researcher and compromises were reached where divergent severities were proposed for some errors. Differing opinions in assessment were mainly caused by students' confusion in attributing an Effective Edition (EE) to a minor severity error (or vice versa) and attributing major errors which were more appropriately considered minor.

All analyses templates for each participant are available at Appendix 5 of this research. A total of 18 outputs were analyzed in this phase, amounting to approximately 4,200 words. The corpus analyzed for EX2 is much reduced when compared toEX1 as, despite using three text chunks instead of one, these were much shorter. Furthermore, fewer participants – one third fewer than EX1 – were tested in this phase.

The total amount of analyzed errors was 318 throughout the five methods. Of these, in all methods, 50.6% (161) were translation errors, and 49.4% (157) were recognition errors. Out of translation errors, 57.1% were minor, 18.6% were major, and 24.3% critical. Out of recognition errors, 63.7% were minor, 17.2% major, and 19.1% critical.

In total throughout the methods and regardless of translation or recognition, 192 minor errors were detected, 57 major, and 69 critical.
As per EX1, translation errors for each method will be discussed first and broken down to the different error typologies to observe their frequency. Translation errors from Method 1 to 3 were, once again, due to either the interlingual respeaker or the simultaneous interpreter's errors, while for Methods 4 and 5 the Machine Translation software was responsible.
Before moving to the delay results, recognition errors are detailed in the following Subsection.

4.1.1 Translation errors

Before moving on to the detail of error typologies and their frequency in the following sections, Figure 40 below shows an overview of the different errors for the five methods. The trend shows content omission errors (in blue) frequently occurring as minor, followed by major errors. Content additions are almost nonexistent, while content substitutions (in gray) confirm the pattern also shown in EX1, namely pertaining almost exclusively to critical severity. Errors of form, both in style and correctness, are confined to minor severity with only one exception.

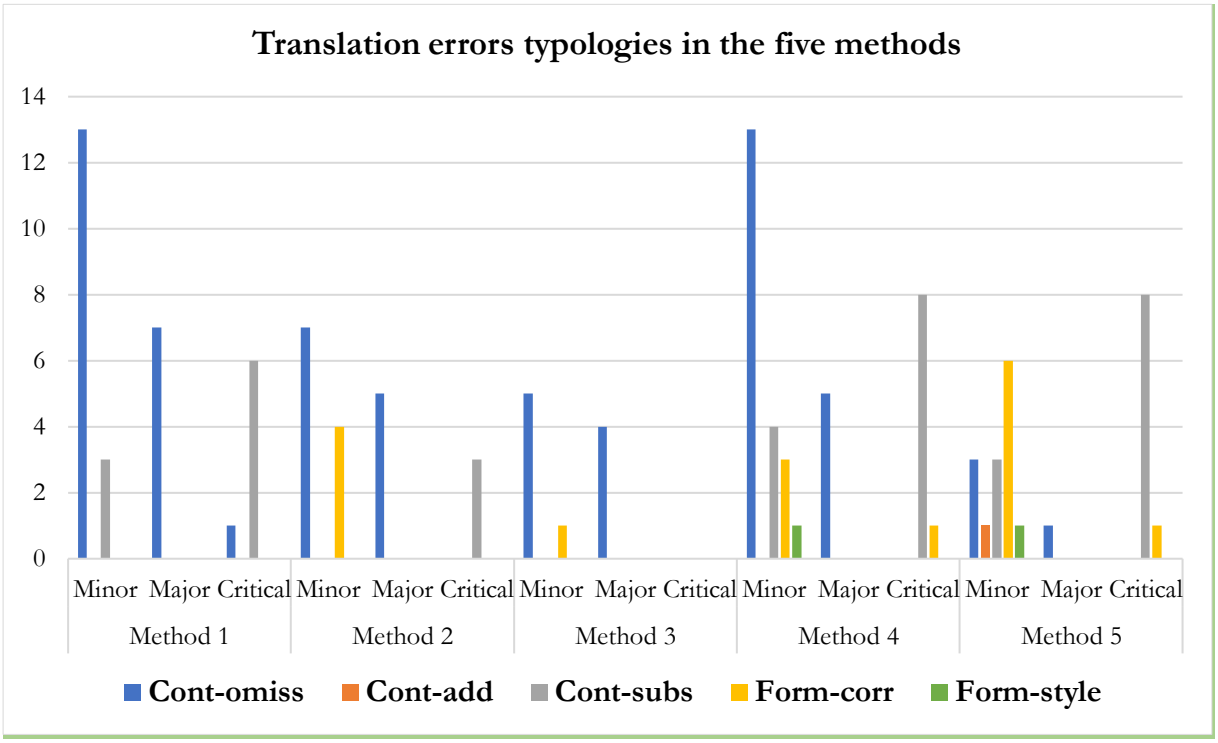


Figure 40 – Overview of the most frequent error typologies according to severity for the five methods

4.1.1.1 Method 1: Interlingual respeaking

Method 1 (interlingual respeaking) was again the most demanding for participants in this version of the experiment. The results of the analyses from this method are still promising since an average of 97.1% was achieved.

Translation errors detected in Method 1 totaled 30. 16 were minor, 7 major, and 7 critical, showing a repeated pattern in all methods whereby minor errors are much more frequent than major, and major more frequent than critical. Concerning the three different text chunks for testing, for the first longer text (speech by the Pope) 13 translation errors were detected, for text 2 (EU Committee speech) there were 10, and for text 3 (speech by president Rajoy) there were 7.

Considering the total number of text errors for this method, all of the 16 minor errors were content based: 13 omissions and 3 substitutions as shown in Figure 41 below, again showing a predominance of omissions in human errors. No additions of content were detected in this case, nor were there any errors of form (in style or correctness). Omissions were the most frequent errors and were mainly caused by the loss of a dependent idea unit, or reiterated ideas.

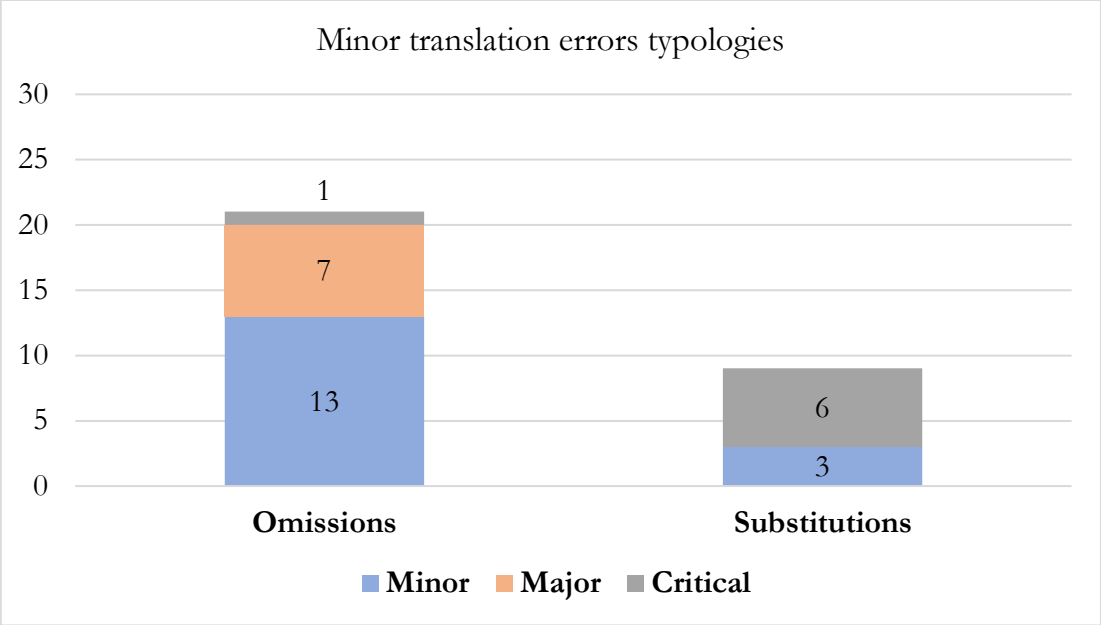


Figure 41 – Minor translation errors typologies for Method 1

All of the 7 detected major errors were of content, and more specifically were omissions thus due to the loss of an independent idea unit by the interlingual respeaker.

There were 7 critical translation errors for this method: as per EX1, no critical errors were due to form and they were all of content: 1 omission, no additions, and 6 substitutions, confirming that critical errors are predominantly caused by substitutions of content.

4.1.1.2 Method 2: SI and intralingual respeaking

Method 2 (SI and intralingual respeaking) functioned much more smoothly in EX2 than EX1. This was probably due to better performance of the participants' personal computers or more stable internet connections, but could also have been due to stronger interpreter outputs and clearer respeaker dictation to the ASR software. Of course, the different language pair Spanish to Italian could have positively impacted results as well. Interestingly, Method 2 ranked first in achieving linguistic accuracy this time, far outperforming the other methods. Since two participants contributed to the final output, errors were subdivided into those caused by the interpreter (INT) (although the analysis sheets do not feature any particular label since, for example, omissions by the interpreter are also automatically attributable to the respeaker who has no access to the ST) and errors caused by the respeaker (RESP).

There were a total of 19 translation errors detected in Method 2 (7 in text 1, another 7 in text 2, and 5 in text 3): 11 were minor (5 done by the interpreter and, consequently, by the respeaker as well, and 6 made only by the respeaker), 5 major (4 by the interpreter and 1 by the respeaker alone), and 3 critical (1 by the interpreter and 2 by the respeaker).

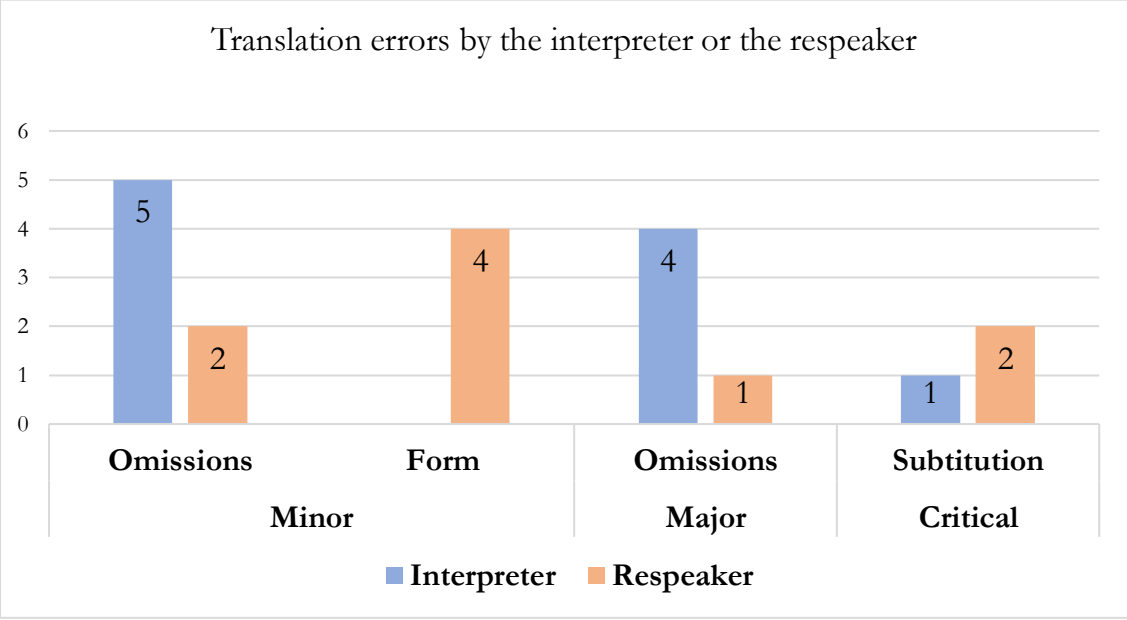


Figure 42 – Overview of translation errors by interpreters or respeakers in Method 2

As shown in Figure 42 above, among the 19 translation errors for this method, 10 were made by the interpreters, and 9 by the respeaker, showing a balance of responsibility between the two participants in the process.

Concerning minor and major severities, many content omission and form correctness errors were detected. Among minor translation errors made by the interpreter (5), all were omissions (therefore, errors of content), while among minor errors by respeakers (6), 2 were omissions and 4 were errors of form correctness due to wrong verb tense or missing gender concordances.

Among major translation errors by the interpreter (4), again 100% were content omission errors, as was the only major error by the respeaker.

Concerning critical translation errors, instead the most frequent error typology was content substitution. The only critical error by the interpreter was, indeed, a substitution, as were the other 2 by the respeaker.

Figure 43 below shows the main different categories according to the single error grading (minor, major or critical), regardless of who was responsible.

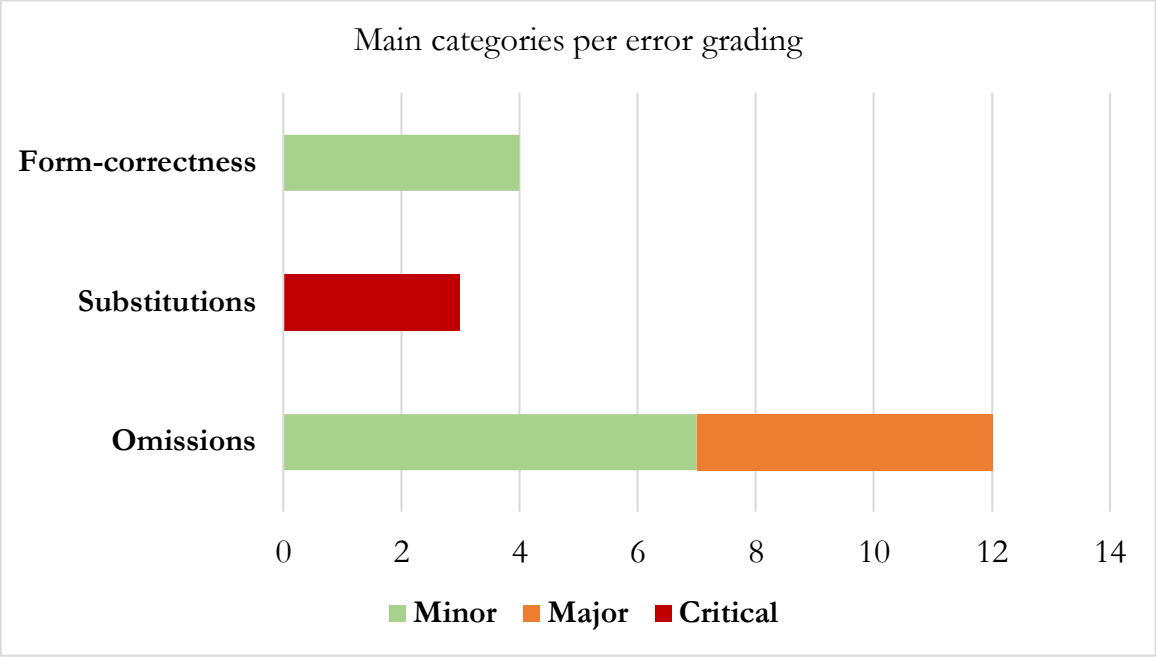


Figure 43 – Main categories per error grading for Method 2

4.1.1.3Method 3: SI and ASR

Method 3 (simultaneous interpreting and automatic speech recognition) for EX2 was carried out using only one ASR machine, Dragon Naturally Speaking.
In the whole method, 298 total translation errors were counted among the three severities: 155 minor, 56 major, and 87 critical. No technical problems were encountered in this phase, since .mp3 recordings were perfectly audible and no interference was captured.

Recognition errors were mostly detected for this method, and only 10 translation errors were detected (2 in text 1, 6 in text 3, and 2 in text 3): 6 minor, 4 major and no critical errors. Among the minors, 5 were of content omission, and the remaining was an error of form correctness. No additions or substitutions were detected. Concerning major translation errors, all were of content and they were all omissions while. No critical errors were identified.

4.1.1.4 Method 4: Intralingual respeaking and MT

Two participants performed in Method 4 instead of one. Since the intralingual respeaking task was performed by non-Spanish native speakers, it was necessary to test the workflows twice and pick the best score out of the two as the outputs could have been influenced by the students' weak proficiency in Spanish as an L2 language, or partially by anxiety. This was only done for this method. Table 32 above features the results for each ST and also for Participant 4, which was then not taken into account for comparison with the other modes. Therefore, translation errors discussed in this section are those made only by Participant 5 (whose results are averaged in Table 33 with the other ones), who scored higher in the performance. Some examples reported below on nonsensical output examples are retrieved by Participant 4 performance as well, as to demonstrate to what extent could the machine misrecognize or mistranslate.

Translation errors in this method are also attributable to the MT system and are labeled '(MT)' errors. Errors by the intralingual respeaker (Spanish to Spanish) cannot be properly deemed as 'translation errors' since the task was intralingual and, therefore, an intralingual analysis through the NER model (see Section 4.2.2, Chapter 2) should have been carried out. Still, errors by the respeaker that should be identified as 'Edition errors' as per the NER model are here labeled as translation errors "(RESP)" in order to isolate omissions, substitutions, or other form errors caused by the human.

Translation errors for Method 4 totaled 35, among which 25 were made by the respeaker, and 10 by the Google Translate MT system (16 in text 1, 12 in text 2, and 7 in text 3).

Concerning the respeaker's translation (edition) errors, 13 were minor, 5 were major, and 7 were critical. Among the minor errors, 11 were of content (8 omissions and 3 substitutions), and 2 of form correctness. All 5 major errors were omissions, and all critical errors were content substitutions.

Concerning the MT errors, 8 were minor and 2 were critical, while no major were detected. Among the minor errors, 6 were of content (5 omissions and 1 substitution), and 2 of form (1 style, 1 correctness), while among the critical errors 1 was a content omission, and 1 of form correctness.

An overview of the different errors' typologies and severities depending on the responsible participant (respeaker or MT) is shown in Figure 44 below.

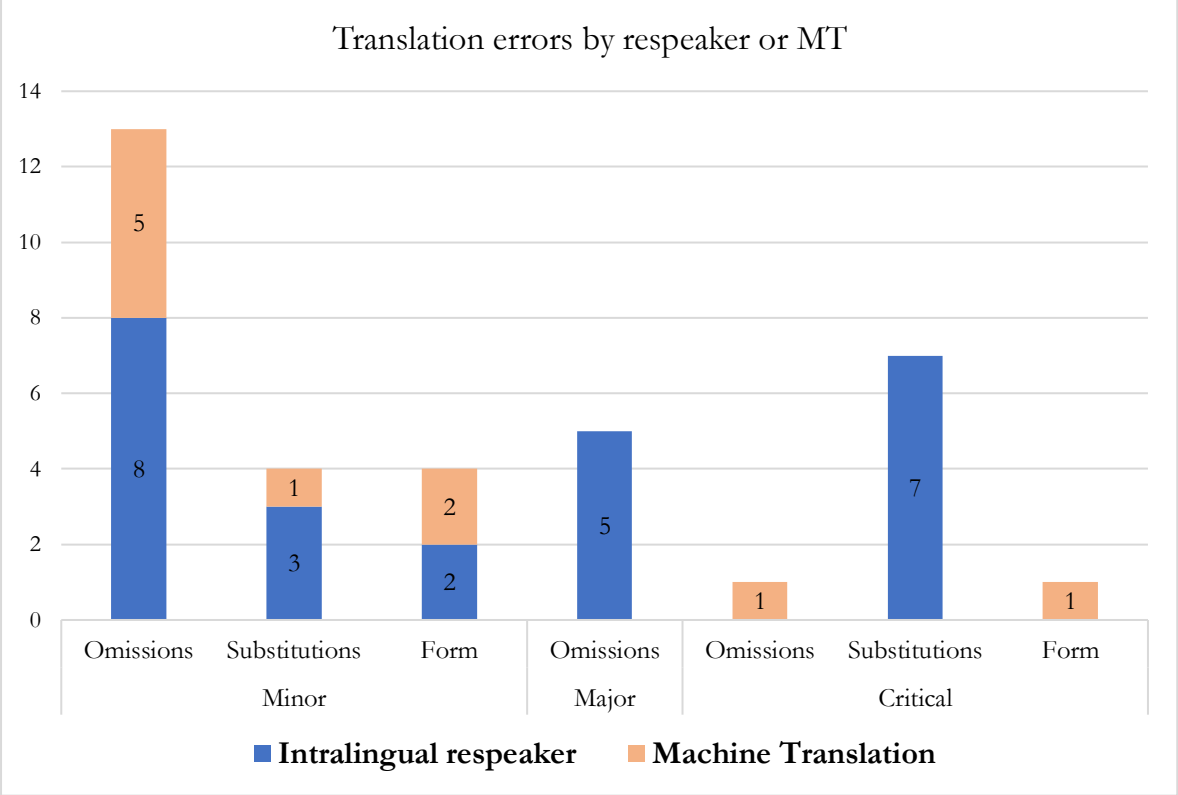


Figure 44 – Overview of translation errors by the respeaker or the MT software in Method 4

The three NTR analyses for the source texts for this method (and the other four) are available at Appendix 5 of this thesis. Nevertheless, some examples are worth noting here. Displayed are the Spanish source text (ST), the Spanish to Spanish respoken text (RT), text recognized by Dragon (ASR), and the final translation by the Google Translate MT system (TT). Given the reduced size of corpus analyzed for EX2, only two examples are displayed here.

Example 1:

ST: "De una crisis no se sale igual. Salimos mejores o peores".

RT: "De una crisis no se sale igual o se sale mejor o peor".

ASR: "La Cristina se sale igua o se sale mejor o peor".

TT: Cristina esce lo stesso o esce meglio o peggio.

Back translation: *Cristina goes out any way either she goes out better or worse*

In this instance taken from text 1 (speech by Pope Francisco) the SR software misrecognized the dictation of 'de una crisis' con 'La Cristina', displaying a sentence that, despite not seeming correct in the given context, could still make sense to the reader to some extent.

Example 2:

ST: "Gracias a todos y de manera muy especial a mi partido sin el cual nada hubiera sido posible".

RT: "Gracias a todos y de manera muy especial a mi partido sin el cual nada había sido posible".

ASR: "Gracias a todos y de manera muy especial a mi partido sin el cual nada había sido posible".

TT: "Grazie a tutti e in modo specialissimo alla mia festa senza la quale nulla sarebbe stato possibile".

Back translation: *Thanks to everybody and in a super special way to my party without her anything would have been possible*

In this example taken from text 3 (speech by President Mariano Rajoy) both the respoken text and the transcription worked well, however in the translation stage there are several aspects to discuss. Firstly, an issue of style: the MT translates "de manera muy especial" as "in modo specialissimo", which in Italian reads informally for a speech given in a parliamentary session. Secondly, the translation of "Partido", which makes sense in the English back translation as a "party" may be both a political party or a celebration, for

example. This is unfortunately not the case for Spanish to Italian, in which it only refers to political parties. "Festa" in Italian is the Spanish equivalent of "fiesta/celebración", which has a completely different meaning to that of the original speech. In addition, despite the text being fed into the MT using the masculine "sin el cual", the software changed it to the feminine following the translation of 'festa'. One positive aspect to note, however, is that when the respeaker poorly dictated a past conditional tense sentence using "había" instead of "habría", the software autocorrected it displaying the correct grammatical structure in the Italian TT.

Example 3:
ST: "En resumen, la pandemia del COVID nos ha enseñado esta interdependencia".
RT: "En resumen la pandemia del COVID nos ha enseñado esta interdependencia".
ASR: "El resumen la bandeja del Covid nos hace ser enseñado esta interdependecia".
TT: "Il riassunto del vassoio Covid ci fa insegnare questa interdipendenza".
Back translation: *The recap of the Covid' tray makes us teach this interdependence.*

Example 4:
ST: "Y vamos a ser más resilientes cuando trabajemos juntos en lugar de hacerlo solos".
RT: "Vamos a ser más resilientes cuando trabajemos juntos en lugar de hacerlo solos".
ASR: "Vamos a ser madres hirientes cuando trabajemos juntos en lugar de hacerlo solos".
TT: "Saremo madri dolorose quando lavoreremo insieme invece che da sole".
Back translation: *We will be grieving mothers when we will work together instead of alone.*

In this last example the misrecognition by the Dragon is critical, since to some extent it can makes sense for the reader. One interesting aspect to notice is that the MT, despite having followed correctly the second half of the sentence and detecting male gender and plural number, consistently to what has wrongly recognized ("madres hirientes") changes it all into the feminine plural.

4.1.1.5 Method 5: ASR and MT

The final tested method for EX2 was fully automated (automatic speech recognition + machine translation) where no human contribution was part of the process. For this mode, the ASR software could not clearly process the speech by only playing the source on the same or another computer, so the same steps as per Method 3 were followed. The source texts were recorded as .mp3 files, and then consequently fed into the SR machine that processed the speech into written text (Spanish to Spanish). The SR written output was then fed to the Google Translate machine translation software (copied and pasted in the translation window), producing the final Italian target text. This was done for the three different source texts.

For this method translation errors are only attributable to the MT and are labeled as "(MT)" errors. Errors by the ASR software in the intralingual process are naturally recognition errors.

Translation errors for Method 5 totaled 24 (16 in text 1, 7 in text 2, and only 1 in text 3). There were 14 minor severity errors: 7 of content (3 omissions, 3 substitutions, and 1 addition), and 7 of form (6 regarded correctness, and 1 stylistic issues). Only 1 major error was detected (content omission), while 9 critical errors were found: 8 content substitutions, and 1 of form correctness due to a mistranslation of the word *ante* translated as "before" (*antes*), instead of "concerning/about" (*en referencia a*).

An overview of the different errors' typologies and severities depending on the responsible party (respeaker or MT) is shown in Figure 45 below.

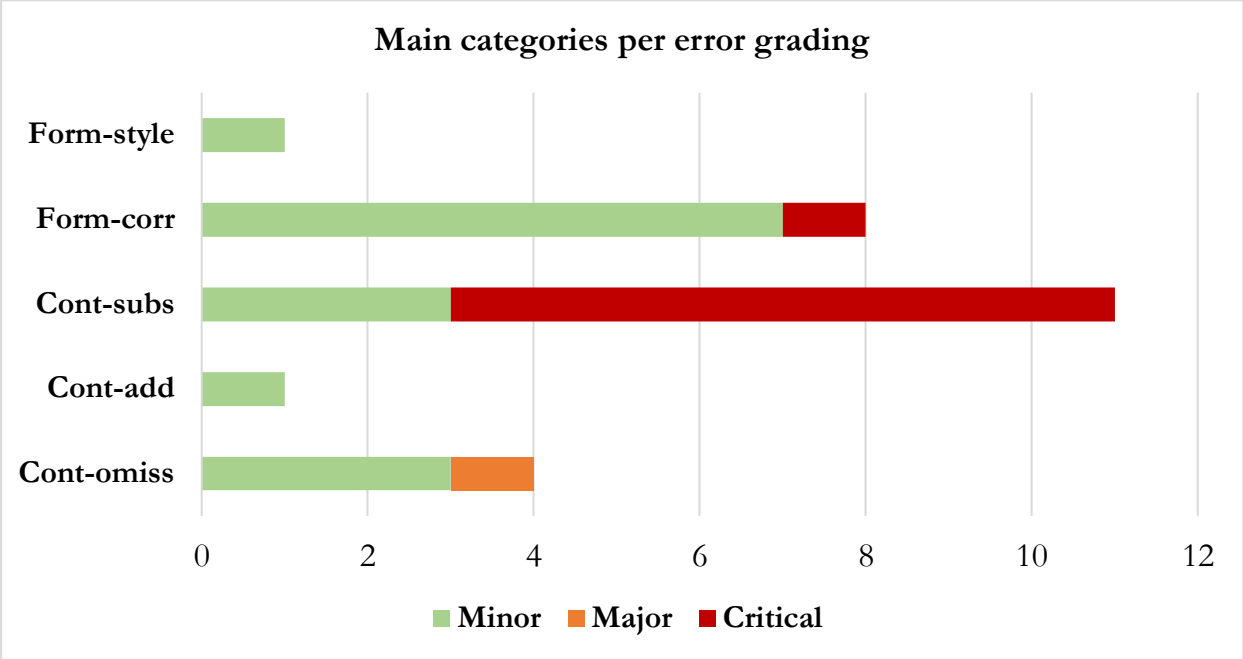


Figure 45 – Main categories per error grading for Method 5

Some outputs in Method 5 are also worth noting, since they are typical examples of situations in which the processing and the final decision on what to display in the subtitles are only attributable to machines with no human monitoring. Examples are provided here with the ST, the text recognized by the software (ASR), and the final target text (TT).

Example 1:

ST: "En resumen, la pandemia del COVID nos ha enseñado esta interdependencia […]".

ASR: "En resumen, la pandemia del obispo nos ha enseñado esta interdependencia […]".

TT: "Insomma, la pandemia del vescovo ci ha insegnato questa interdipendenza […]".

Back translation: *To sum up, the bishop pandemic taught us this interdependence.*

This first example is taken from text 1, and the word 'Covid' is misrecognized as 'obispo' (bishop). While it would probably be clear that a 'bishop pandemic' is not the intended meaning, in this case it is the Pope speaking in the video so there is some link between that

and the topic of bishops. Either way, what can be concluded is that displaying a target text as in the example above can vary in results according to the situation.

Example 2:

ST: "Si alguien se ha sentido, en esta cámara o fuera de ella, ofendido o perjudicado le pido disculpas. Gracias a todos".

ASR: "Si alguien se ha sentido en esta cámara oscura, bello, ofendido o perjudicado le pido gracias a todos".

TT: "Se qualcuno si è sentito bello, offeso o leso in questa camera oscura, chiedo grazie a tutti".

Back translation: *If someone felt good looking, offended, or damaged in this darkroom, I ask thanks to everyone.*

There are several problems in this second example taken from the speech by Mariano Rajoy. Firstly, "cámara" (referring to the Parliament) is misrecognized as "darkroom" – as in for photography, which makes very little sense, or even as in a 'shady room', which could be particularly misleading given the diplomatic scenario in which the speech is given. Secondly, part of the expression "fuera de ella" is omitted and only 'good looking' (bello) is recognized, which makes no sense in this context to the extent that it sounds ridiculous to the reader and, thus, is recognized as a mistake. Lastly, the ST "disculpas" is omitted, which is an exception here since the machine rarely completely misses parts of speech. Together with the misrecognition of the full stop at the end of the sentence, the beginning of the new one is linked to the previous, resulting in misinformation: "pedir gracias" should be "decir gracias" and does not sound natural but could be seen as a stylistic error and not as a double omission (of "disculpas" and of the punctuation mark). Here the subtitling displays a worrying mistranslation.

Sometimes, Method 4 also featured debatable translation choice at a lexical level, as in the case of "clima" was translated as "meteo" (weather), instead of "climate", since reference was made to climate change, of course.

4.1.2 Effective editions

Table 34 below shows all of the effective editions (EEs) for each participant according to which method they were involved in. Method 2 differentiates between EEs made by the interpreter (INT) and the respeaker (RESP) in the intralingual respeaking task. Among the three used for testing, some source texts more than others featured room for reformulation (text 1 especially, since many repetitions were present).

Method	Participant	Text	Responsible		Total EEs
Method 1 – interlingual respeakers	Participant 1	Text 1	/	10	**15**
		Text 2		2	
		Text 2		3	
Method 2 – SI + intralingual respeaking	Participant 3	Text 1	RESP	7	**12**
			INT	5	
		Text 2	RESP	0	**2**
			INT	2	
		Text 3	RESP	0	**1**
			INT	1	
Method 3 – SI + ASR	Participant 2	Text 1	Dragon	5	**8**
		Text 2	Dragon	2	
		Text 3	Dragon	1	
Method 4 – intralingual respeaking + MT	Participant 5	Text 1	/	3	**4**
		Text 2		0	
		Text 3		1	
Method 5 – ASR + MT	NA	/	/	/	**/**

Table 24 – Effective Editions (EEs) per participant in the five methods

4.1.3 Recognition errors

Recognition errors for the five tested methods were attributable to the ASR software; a total of 157 errors were detected.

In Method 1, 9 recognition errors were found (7 in text 1, one in text 2, and one in text 3): 5 were minor, 3 were major, and 1 was critical.

In Method 2, 6 recognition errors were found in the three texts (3 in both text 1 and text 3, and none in text 2: 5 were minor, 1 major, and none critical.

Method 3 registered a higher number of recognition errors: 57 in total were assessed (compared to the 10 translation errors), divided as follows: 51 minor, 2 major, 4 critical. This method did not allow for human monitoring and edition since the interpreter's output was taken directly by the machine.

In Method 4 there were 7 recognition errors (6 in text 1, only 1 in text 2, and none in text 3): 3 were of minor severity, none were major, and 1 was critical.

For Method 5, as per Method 3, many more recognition errors were detected since there was no human intervention in either dictation to the SR software or monitoring of the written output, as can be seen in Figure 46 below. A total of 61 recognition errors were detected for the fully automated mode (34 in text 1, 11 in text 2 and 16 in text 3), surpassing the errors counted in Method 3 by 10. Of all errors, 36 were minor, 12 were major, and 13 were critical.

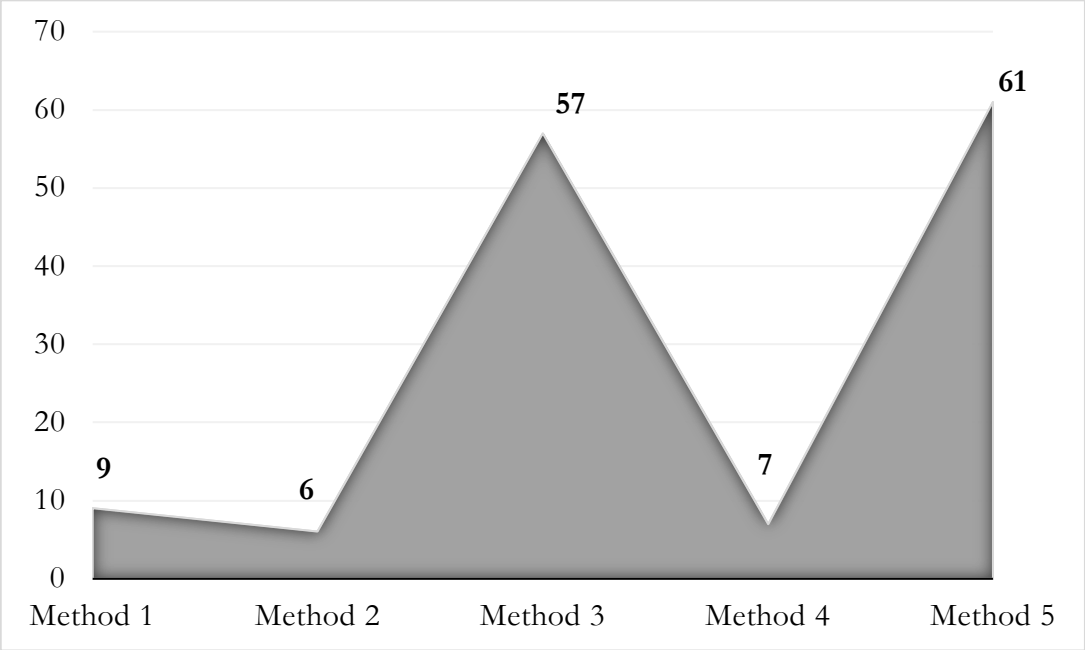


Figure 46 – Recognition errors in the five methods

4.1.3.1 Punctuation errors

The major issue during speech recognition was punctuation marks. It is too difficult for the machine to detect the ST speaker's intonation and place full stops, commas, question marks, etc. in the right positions. As a result, in Methods 3 (SI plus ASR) and 5 (ASR plus MT), in which punctuation was added automatically by the speech recognition software, punctuation is in some parts nonexistent or is not placed where it belongs for proper readability of the subtitles.

For Method 3, among the 51 minor recognition errors approximately 80% were punctuation based (41 in total). No major error was caused by misrecognition of punctuation marks, while out of the 4 critical errors, 2 were as a result of punctuation.
For Method 5, among the 36 minor recognition errors approximately 73% were due to a lack of punctuation (26 in total), while out of 13 critical errors, only 1 was due to punctuation.

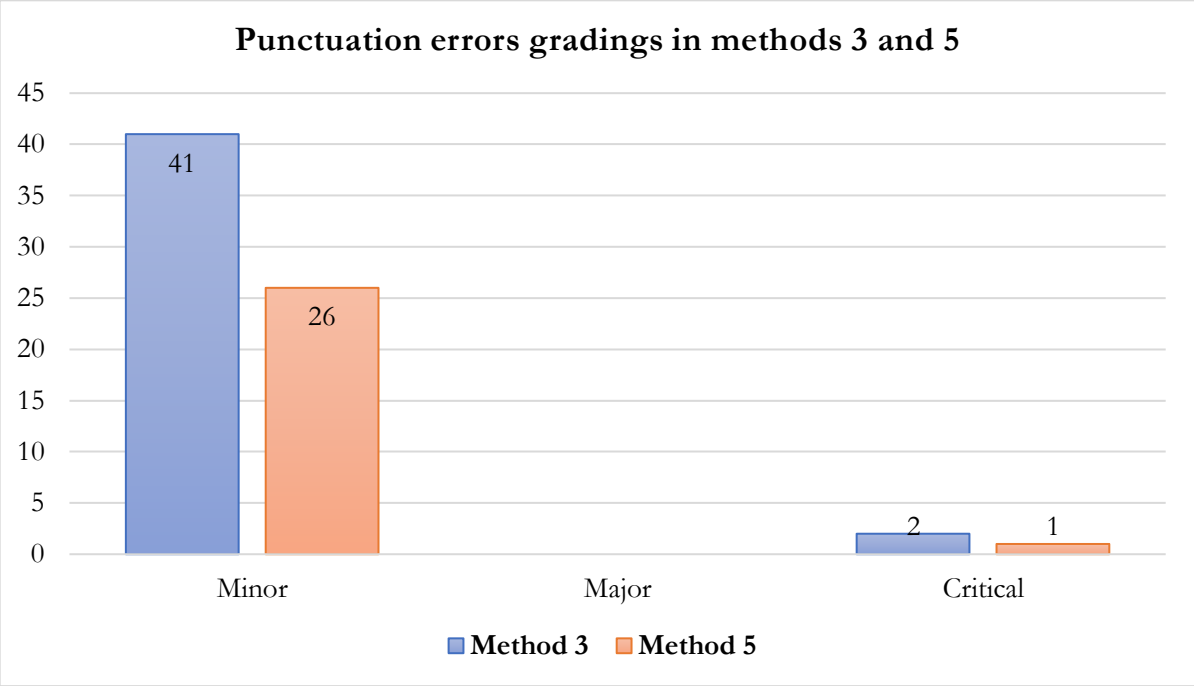


Figure 47 – Punctuation marks error gradings in Methods 3 and 5

The three critical errors that were due to punctuation appeared in Method 3 (examples 1 and 2) and Method 5 (example 3).

Example 1:

ST: "Es el momento de actuar, estamos en el límite. Quisiera repetir un dicho viejo español".

TT: "È il momento di agire perché siamo al limite secondo un vecchio detto spagnolo".

Back translation: *It is the moment to act because we have no time left according to an old Spanish saying.*

Example 2:

ST: "Señoras y señores diputados, seré muy breve. A la vista de lo que todos sabemos […]".

TT: "Signore, signori deputati, cercherò di essere molto breve come tutti sappiamo […]".

Back translation: *Ladies and gentlemen MPs, I will try to be concise as we all know.*

Example 3:

ST: "[…] es bueno recordar cosas que nos decimos mutuamente para que no caigan en el olvido. Desde hace tiempo estamos tomando más conciencia […]".

TT: "è sempre bene ricordare le cose che ci diciamo nell'oblio di quel tempo stiamo diventando più consapevoli…".

Back translation: *It is always good to remember things that we say in the oblivion of that time we are becoming more and more aware.*

In this last example in particular, the sentence does not flow and it is clear that there was a problem ('oblivion'). The lack of a full stop makes it seem as though 'desde hace tiempo' refers to the previous sentence, thus giving it a different meaning.

4.2 Delay

Similarly to EX1, in EX2 delay was calculated by choosing one sentence ending per minute in the ST and assessing the delay of that sentence ending compared to when the text first appeared on screen. For Methods 3, 4 and 5, delay calculations were speculative since the whole process was split into different parts (interpretation and feed into the ASR software for Method 3, and intralingual respoken text copied and pasted in the MT window for Methods 4 and 5).

For Method 3, it was only possible to calculate delay for the first part, namely the *décalage* of the simultaneous interpreter while translating, and then an average of 2 extra seconds was added as per the time the ASR software usually takes to process the audio input and produce the written text. For Method 4, the intralingual respeaker delay was calculated and then 1 extra second was added to emulate the time the MT software usually takes to produce the translated output. Lastly, for Method 5, two seconds for the ASR and one extra second for the MT were allowed.

Table 35 below shows the delay calculated in each method for each source text, also offering an average among the three calculations. Participant 4 performing in Method 4 is displayed here since the calculations were carried out for this output as well, but it is highlighted in gray since, in the end, it was not considered for the final average. Only the best outcome among the two participants for Method 4 was taken into account as per linguistic accuracy results.

Methods	**Methods**	**Delay average per text (seconds)**	**Delay average per method (seconds)**
Method 1 (Participant 1)	Text 1	7.2	7.0
	Text 2	8.3	
	Text 3	5.5	
Method 2 (Participant 2)	Text 1	7.6	11.3
	Text 2	10.7	
	Text 3	15.5	
Method 3 (Participant 3)	Text 1	4.2	5.7
	Text 2	5.3	
	Text 3	4.5	
Method 4 (I) (Participant 4)	Text 1	5.6	[5.0]
	Text 2	3.0	
	Text 3	6.5	
Method 4 (II) (Participant 5)	Text 1	7.4	6.8
	Text 2	6.0	
	Text 3	7.0	
Method 5 (Participant 6)	Text 2	3.0	3.0
	Text 3	3.0	
	Text	3.0	

Table 35 – Participants' delay results in seconds[39]

[39] Results in squared brackets by Participant 4 are not taken into account in the delay average per method shown in Table 36.

Table 36 displays the delay average for each method more clearly, only including Participant 5 outputs for Method 4 this time. Together with the average per method, the ranking is displayed, where fully and more-automated methods were shown to be faster than human-centered ones in processing both intralingual SR and automatic translation.

Methods	Delay average (seconds)	Rank
Method 1	7.0	4
Method 2	11.3	5
Method 3	5.7	2
Method 4	6.8	3
Method 5	3.0	1

Table 36 – Delay average per method

4.3 Cost

The present research did not seek to quantify costs for the different ILS methods. Nevertheless, the experiments conducted by the University of Vigo (Romero-Fresco & Alonso-Bacigalupe, 2022) and by the University of Surrey (Dawson, 2021) did offer an overview of potential speculative costs for the implementation of each method considering the number of professionals and machines involved. Drawing upon their findings, Figure 48 below shows the expected cost of the service depending on the length of the task. For assignments shorter than 30 minutes, for instance, in Methods 1, 2, 3 and 4 the number of humans involved is the one relevant to the mode itself (i.e., 2 human professionals for Method 2: an interpreter and an intralingual respeaker). In the third column, the number of humans that would be needed for assignments exceeding 30 minutes in length is

displayed: as simultaneous interpreters typically working in pairs in the booth, human professionals in each role would require another colleague to take turns with all the modes. The longer the task, the more professionals would be required and, therefore, the higher the cost. The advantage of machines is that, regardless of the length of the assignment or live event, the cost remains the same. Table 48 below shows the five methods ranked from the cheapest one (1) and the most expensive (5) in terms of budget.

COST			
Modes	**<30 minutes**	**>30 minutes**	**Rank**
1. Interlingual respeaking	1 human	2 humans	4
2. Simultaneous interpreting + Intralingual respeaking	2 humans	4 humans	5
3. Simultaneous interpreting + Automatic Speech Recognition	1 human + machine	2 humans + machine	3
4. Intralingual respeaking + Machine Translation	1 human + machine	1 human + machine	2
5. Automatic Speech Recognition + Machine Translation	2 machines	2 machines	1

Figure 48 – Potential cost of each method (Romero-Fresco & Alonso-Bacigalupe, 2022)

4.4 Final thoughts

EX2 tested five methods of producing live interlingual subtitling from Spanish to Italian, adding to the first three also testing intralingual respeaking in EX1 (Spanish to Spanish) plus MT, and ASR plus MT. Fewer participants were tested which therefore did not allow for a comparison between more teams, but this time better scores were obtained on average thanks to the fewer technical issues encountered during the experiment. One

major improvement was detected in the intralingual respeakers' dictation to the ASR software which was optimal in some cases and contributed to higher accuracy. Interestingly for this second workshop edition, dictation to the SR software was introduced sooner when compared to the first edition of the workshop (see Section 3, Chapter 4.3), which may show that it helps trainees to have more practice of this in order to better perform in the respeaking tasks.

Method 2 (SI plus intralingual respeaking) achieved the highest level of accuracy thanks to good interpretation and good intralingual respeaking from Italian to Italian. Method 3 (SI plus ASR) ranked second together with Method 4 (intralingual respeaking plus MT), with almost no difference from Method 1 in interlingual respeaking (the first two scoring 97.2%, the third 97.1%), although none of them reached the required threshold of 98% on average. The three methods do seem promising, with Method 3 having scored 98% in text 2, Method 4 98.5% in text 3, and Method 1 97.9% in text 3, for example. Method 5 ranked bottom with a much lower NTR score for accuracy.

Concerning delay, an inverse correlation with accuracy was detected: those methods that were faster in production of subtitles were the ones that scored lower accuracy rates, and those that required a longer delay were more accurate. In particular, Method 2, which scored the highest in linguistic accuracy, was the slowest since two participants were involved in the process. Referring back to Romero-Fresco's citation (see Chapter 2, Section 4.1), the correlation between accuracy and delay is all the more important and intended as a 'trade-off': more accurate subtitles take more time to be produced (and corrected), while less accurate ones are cued with much shorter delay.

5. Discussion

EX1 and EX2 can only be partially compared since there are differences in the methodologies. For EX1, only one video which was longer than any other used in EX2 was tested, featuring almost no specialized terminology, and focused on testing and comparing the first three methods. For EX2, a different language combination was considered, and three shorter videos were used for testing, one of which (text 2) did present some specific terminology. Furthermore, EX1 was carried out with 15 participants divided into 5 teams of 3, while EX2 only involved 5 participants, thus no comparison between different teams' performance could be carried out.

Concerning language pairs, it is also possible that the Spanish to Italian testing scored better results overall due to the similar syntactical structure of both languages. However, it is interesting to stress how this leads to more frequent calques by respeakers and interpreters, but also by the machine (as it was shown in some examples when the translation was carried out by the MT system). In such cases, and in particular with texts with specific terminology, the (human) linguistic and strategic competence can help in avoiding claques and finding suitable translation options more than the term knowledge itself (Pagano, 2020b), something that still remains to research for more automated methods, especially Method 5.

In general, some differences can be detected when comparing results from the first three methods of the two experiments for this research, but they also show some consistencies. While delay in the production of the subtitles coincide almost to the second in both cases, accuracy scores differed. In noting the higher scores in two methods for EX2, it is important to remember that only one participant performed in each method, not 5 as in EX1. Having more individual performances can easily lead to a lowering of the average, especially given the participants were students and not professionals. Furthermore, many more technical problems were encountered in EX1 that partially hindered performance.

Regarding accuracy, as shown in Table 37, the first difference can be seen in the higher accuracy score reached in EX2 with **Method 2** (SI plus intralingual respeaking). Among all the outputs analyzed throughout the research, this Method in EX2 reached the highest percentage using the NTR mode: if in EX1 the accuracy was poor (96.9% on average), with

EX2 98.6% was accomplished (very good accuracy). In both tests however, the delay in this workflow is the longest (up to 11.4 seconds) since two professionals are involved, thus time is extended. Using two professionals for the assignment also signifies higher costs.

Method 3 also produced better results in EX2 than EX1 thanks to a very good interpretation output, and good dictation by the intralingual respeaker. In addition, the source texts – despite having three of them tackling different topics – were much shorter and probably less tiring for the participants. For EX1, the AppTek SR software scored higher than Dragon averaging 95.9% which was insufficient. In EX2, this mode scored as much as Method 1 and Method 4 (97.2%), which is still considered poor for accuracy but still improved and promising. The major problem for this workflow in both experiments was punctuation: it was almost nonexistent, and where commas and full stops were present, they were misplaced and did not allow for clear reading flow. This can be due to the typical non-natural prosody of the simultaneous interpreters that, fed into the SR system, did not allow for proper recognition of pauses and intonations (prosody itself). Delay in both cases was acceptable (approximately 5.7 seconds). Since only one professional is required for this mode, cost is again expected to be low.

Concerning **Method 1**, the results were consistent in both experiments, reaching 97.2% and 97.1%, which is not high but still promising. Delay for this mode is around 7 seconds on average, which is acceptable especially for interlingual live subtitling. Furthermore, the cost is contained since only one professional is required for the task despite the skills required and high level of technicality.

The last two methods were exclusively tested as part of EX2 for this research.
Method 4 scored 97.2%, similarly to Methods 3 and 1, thus not reaching the required threshold but demonstrating encouraging results. It is worth mentioning that, while for the other methods that scored the same accuracy rate, omissions were a major problem, for Method 4 the biggest issues were some automatic translation solutions that made no sense. In this respect, it is important to underline that, depending on the situation, some mistranslations occur and can create misunderstandings. Another major point is given by

the frequent content omissions by the respeaker, which could be avoided by Method 5, in which presumably omissions are less likely to occur. While punctuation was good as it was dictated by the intralingual respeaker from Spanish to Spanish, the TT displayed some disrupted text and some information that was not monitored and edited by any human, which could result in issues for technical or formal settings for example. Still, Method 4 demonstrated good potential, given that it is very fast and has a very low cost.

Method 5 scored the least in accuracy, even when compared to EX1 outputs (95.4%). A major issue for this mode was that it was fully automated, thus no human contribution was given either in the intralingual or in the interlingual process, and not even in the monitoring phase. This led to some major misrecognitions from Spanish to Spanish that were then transposed to Italian by the machine translation, displaying some subtitles that did not make any sense or, worse, created a new and misleading meaning to the text. Also, punctuation marks were again a major issue in this mode since they were added automatically by the ASR software and were then held in place during the translation phase. As per Method 3, punctuation was almost nonexistent or misplaced, therefore hindering smooth readability of the TT. A fully automated method is likely to omit very little from the ST, which is an important advantage. Nevertheless, in the example reported of the missing word *disculpas*, there is a chance of creating misunderstandings through omissions even with this workflow. Despite this, further investigation is needed on a much wider corpus of outputs: example 2 reported on text 3 of the research (Section 4.1.1.5) demonstrates that if the source audio input is not cleared from all sound interference and background noise, it becomes difficult for the ASR to grasp precisely what is being said.

Both Method 4 and Method 5 are still to be validated for specialized source texts (Romero-Fresco & Alonso-Bacigalupe, forthcoming). Examples like the ones provided for the translation of "clima" (which occurred in both methods since due to MT) or *partido* (discussed above) hint that they are likely to fail in texts in which specific terminology is used.

The three and five methods tested in EX1 and EX2 can also be compared with the results obtained in the experiments by the University of Surrey and by the University of Vigo to draw some conclusions. Before doing so, some specifications on methods and settings for both experiments are summarized. Taking into account that the University of Surrey (Dawson, 2020) tested from Spanish to English and that the University of Vigo (Romero-Fresco & Alonso-Bacigalupe, 2022) from English to Spanish, they both investigated the same five methods (or modes) of this research, through one only experiment that included:

- Mode 1. Interlingual respeaking: a single STT interpreter

- Mode 2. Simultaneous interpreting + intralingual respeaking: a simultaneous interpreter who translates the source text and an intralingual respeaker

- Mode 3. Simultaneous interpreting + ASR: a simultaneous interpreter who translates the source and an ASR engine

- Mode 4. Intralingual respeaking + MT: an intralingual respeaker and a MT engine

- Mode 5. ASR + MT: an ASR engine and a MT engine

Materials used by Romero-Fresco and Alonso-Bacigalupe were two videos: the first TEDx Stockholm talk by Greta Thunberg on 'School strike for climate – Save the world by changing the rules' used by the researcher in her EX1, and the speech 'Readers are leaders' by Phuong Anh Nguyen Ngoc for another TEDx event in Hanoi (Vietnam). This last ST was a little longer since it lasted 15 minutes and accounted to 2.537 words delivered with a 165 wpm rate.

Materials used by Dawson were other two TED talk videos in Spanish: 'Cómo nos manipulan en las redes sociales' (that lasted 10:50 minutes with a speech rate of 104 wpm), and 'Educar en el feminismo' (12:00 minutes, 133 wpm).

The main substantial difference between this research and the other two concerns participants, since in this case only students were involved in the experiments, while for Vigo and Surrey were professionals in different fields with long prior experience: two

professional interlingual respeakers, two simultaneous interpreters, two intralingual respeakers (in the case of Vigo, two for Spanish>Spanish and two for English>English respectively). As far as machine were concerned, Surrey used Dragon Professional v15 as ASR engine, while Vigo AppTek, whereas both used Google translate as MT software.

The results extracted from the experiment by the University of Vigo show that most accurate modes are 2 (SI + intralingual respeaking ES>ES) and 4 (intralingual respeaking EN>EN + MT) with 98.8% threshold in the NTR calculation; mode 1 (interlingual respeaking) obtains good results too (98.4%), while modes 3 (SI + ASR) and 5 (ASR + MT) score lowest (97.4% and 97.2%, respectively).
The results extracted from the experiment by the University of Surrey did not reach in any method the 98% threshold, as per this research EX1, showing mode 1 (interlingual respeaking) ranking first with 97.9%, followed by mode 2 (SI + intralingual respeaking EN>EN), mode 4 (intralingual respeaking ES>ES + MT), mode 5 (ASR + MT) and, last, mode 3 (SI + ASR), confirming Vigo's pattern.

Concerning delay, for this research 2 seconds more on average for each method were detected, but very similar results were obtained overall and the trend of 'the more accurate, the slower', 'the least accurate, the fastest' was confirmed. For the sake of clarity, delay between 4 to 6 seconds is still considered as 'acceptable', and 'long' is when it reaches 8 to 12 seconds instead.

At this stage, a comparison between results obtained in EX1 and EX2 of this research is outlined in Table 37 below, followed by Table 38 showing the same review for the University of Vigo's results, and Table 39 for the University of Surrey's findings.

Method	Accuracy EX1	Accuracy EX 2	Delay EX1	Delay EX2	Cost
Interlingual respeaking	Poor	Poor	Acceptable	Acceptable	Low
SI + intralingual respeaking	Poor	Good	Long	Long	High
SI + ASR	Insufficient	Poor	Acceptable	Acceptable	Low
Intralingual resp. + MT	/	Poor	/	Acceptable	Low
ASR + MT	/	Insufficient	/	Short	Very low

Table 37 – Comparison between EX1 and EX2 results[40]

Method	Accuracy rate	Delay	Cost
Interlingual respeaking	Good	Acceptable	Medium high
SI + intralingual respeaking	Good	Long	High
SI + ASR	Poor	Acceptable	Low
Intralingual resp. + MT	Good	Acceptable	Low
ASR + MT	Poor	Short	Very low

Table 38 – Results by the University of Vigo (Romero-Fresco & Alonso, 2022)

Method	Accuracy rate	Delay	Cost
Interlingual respeaking	Almost acceptable	Short	Low
SI + intralingual respeaking	Almost acceptable	Almost acceptable	High
SI + ASR	Insufficient	Short	Low
Intralingual resp. + MT	Poor	Short	Low
ASR + MT	Insufficient	Short	Very low

Table 39 – Results by the University of Surrey (Dawson, 2021)

40 Refer to Table 2, page 76 for Classification of performances in reference to the NTR model.

Despite some differences in the methodologies (fore and foremost the level of expertise in participants and tools), and despite some differences in NTR scoring, the results of the three research projects can be considered consistent with one other, regardless of language pairs. For EX1 and EX2 of this research, the accuracy levels were lower than those obtained by Romero-Fresco and Alonso-Bacigalupe, and slightly lower but more consistent with those obtained by Dawson, which is possible as they both used professionals for the testing (interpreters and respeakers), while for this research only students were involved with lower proficiency from the beginning.

It is worth noticing that in Vigo's results Method 2 scored better than Method 1, while for both this research and for Surrey, Method 1 always performed better than mode 2. Interestingly, Method 4 obtained better results for Vigo, while ranked lower in the other two experiments. Nevertheless, the most accurate methods were in the three cases the ones involving at least one human participant, while the least accurate was Method 5 in both this and Vigo's research, ranking only fourth in the University of Surrey's testing.

5.1 Educating simultaneous interpreters for accessibility

In light of what has emerged observing the results from EX1 and EX2, as well as from the other parallel experiments, the four methods that involve at least one human professional in the ILS process deliver the highest accuracy and an acceptable delay in the broadcasting: thus good quality subtitles. The fully automated method is still lagging behind, although both ASR and MT systems are likely to improve very much in the coming years, becoming even more precise. For some methods more than others, it seems fundamental to better train the participants in some of the specific features that help with a smoother process and higher quality outcome. First and foremost, it would seem very useful to include training in respeaking for students undergoing an academic degree in the field of Translation and Interpreting. Their background in simultaneous interpreting, translation competences and, in some cases, subtitling skills can be easily exploited in educating students for Media Accessibility. If trained on how to interact with the SR software

(dictation, punctuation, segmentation above all) simultaneous interpreters have great potential in perfecting their interlingual respeaking skills and performing these kinds of tasks. If trained to shorten their *décalage*, using Method 2 they could deliver subtitles with shorter delay, thus scaling up the ranking even more. Finally, if trained on how to cleanse their outputs of oral hindrances such as hesitations, reformulations on the go, false starts, and so on, ASR systems could process their audio inputs more easily in the speech-to-text process. This could surely happen with a substantial improvement in SR software of automatically added punctuation. Concerning Method 4, intralingual respeakers could simplify their dictation so that it might be more easily translated by the MT.

Conclusions

1. Introduction

The final part of this thesis presents the conclusions of the doctoral research. Firstly, the chapter provides the answers found to the RQs outlined in the introduction. Some useful considerations and findings discovered during the investigation, though not included in the RQs, will also be presented. Secondly, bias and limitations of the study and the experiments are discussed, proposing possible improvements for future works. The thesis ends with interesting outlets and possible developments for future research in this field, and some final thoughts deriving from the discussion in Chapter 5.

2. Answers to RQs

The research in its entirety was designed around the first main RQ:

How do different ILS methods compare in terms of quality, when quality is referred to as the joint result of both linguistic accuracy and delay in broadcasting, for English and Spanish to Italian language combinations?

Two experiments were carried out, EX1 from English to Italian testing three methods (interlingual respeaking; SI + intralingual respeaking; SI plus ASR), and EX2 from Spanish to Italian testing five methods (interlingual respeaking; SI plus intralingual respeaking; SI plus ASR; intralingual respeaking plus MT; ASR plus MT), each contributing both similarities and some differences in the attempt to answer this first main RQ question. Both

experiments gave birth to four sub-RQs that are answered before informing the main one, and are classified under EX1 (RQ1, RQ2, RQ3, and RQ4), and EX2 (RQ1, RQ2, RQ3, and RQ4). Throughout the research, some qualitative data was also collected, mainly concerning training in interlingual respeaking, and aimed to shed light on the following additional RQs:

RQ5: Is a master's degree in Translation and Interpreting an adequate environment to train students in intra and interlingual respeaking?

RQ6: Can respeaking training be useful in better educating students in Media Accessibility and, specifically, in DHOH accessibility needs?

In drawing the conclusions for the thesis, the four sub-RQs for both experiments will be answered first, since they contributed to the conclusion of the main RQ itself, which will be answered at the end. Then, RQ5 and RQ6 will be answered together with some insights that contribute to the relevant subject matter tackled in the research, although not included in the specific RQs.

2.1 First experiment – RQ1, RQ2, RQ3, RQ4

EX1 – RQ1:

Among the three considered methods, which delivers the highest linguistic accuracy?

In the first experiment, three methods of producing ILS were tested and compared. Analysis of the performances found that Method 1 delivered the highest scores in linguistic accuracy through the NTR model, having reached 97.2%. Method 2 followed with an average of 96.9, and Method 3 ranked bottom, where the AppTek ASR software delivered higher accuracy than Dragon Naturally Speaking, reaching 95.9%.

EX1 – RQ2:

Among the three considered methods, which one is broadcast with minimum delay?

Method 3 was found to be the fastest. This was the mode in which ILS workflow was delivered split into two tasks: the interlingual task carried out by a human (interpreter) and the intralingual task by the ASR machine, without editing. Methods 1 and 2 followed, with the latter taking longer because two (human) participants were involved, thus delay comprised of both the interpreter's *décalage*, and the intralingual respeaker's output.

EX1 – RQ3:

Is there, among the three considered methods, one that can provide higher-overall quality ILS, taking into account both accuracy and delay as the main factors for evaluation?

By combining linguistic accuracy and delay results for each method, an inverse correlation can be observed as pointed out in the EX1 discussion in Chapter 5: the least accurate method (Method 3) is the one fastest in delivering subtitles, while the more accurate Methods 1 and 2 take longer, especially Method 2 (reaching up to 11 seconds of delay). Therefore, the answer to RQ3 can be found again in the trade-off between accuracy and delay, and in the industry's needs and standard requirements. The 'higher overall quality' that RQ3 sought to find is then determined by the setting and the situation in which the ILS is delivered: the need for faster subtitles is to the detriment of accurate transcription and translation, or better accuracy but provided slowly. As per Eugeni (2020), delay tends to be considered secondary to accuracy, especially given that the delay for Method 1 (7 seconds on average) was perfectly acceptable. To conclude, in EX1 it is Method 1

(interlingual respeaking) that could by many means be considered the method that provides higher quality, followed by Method 2, and Method 3 (AppTek).

EX1 – RQ4:

Are there other variables to consider in searching for higher-quality methods for ILS, and what are they?

From the preliminary answer provided to RQ3, it is indeed possible to observe that there are other important variables to consider in searching for higher-quality ILS methods. Among these, the formality versus informality of the setting in which subtitles are to be displayed: given the high occurrence of typos, misspellings and unintelligible words displayed especially using Method 3, it would not be advisable to use the method in formal situations such as institutional conferences, work meetings, educational environments and so forth. The highly automated Method 3, in which the final say on what and how to broadcast is left to the machine with no monitoring of, editing, or correcting the output can be hazardous. Another important factor to ponder would be that related to punctuation marks. Both speech recognition software added punctuation automatically starting from pauses and intonation by the speaker. Unfortunately, none of the machines succeeded in placing full stops in the right position, thus making it difficult for audiences to follow a text that does not feature proper punctuation. Last but not least, the cost is an important factor to consider when choosing the best-suited workflow depending on the available budget.

2.2 Second experiment – RQ1, RQ2, RQ3, RQ4

EX2 – RQ1:

Among the five considered methods, which one delivers the highest linguistic accuracy?

In the second experiment, five methods of producing ILS were tested and compared. The analysis of the performances found that Method 2 delivered the highest scores in linguistic accuracy through the NTR model, reaching an average of 98.6%. Afterwards followed methods 3, 4 and 1 with almost equivalent scores (97.2% or 97.1%), with Method 5 ranking bottom.

EX2 – RQ2:

Among the five considered methods, which one is broadcast with minimum delay?

Method 5 was by far the fastest. This was indeed the mode in which the ILS workflow was delivered entirely automatically by machine, hence the performance was much better in terms of speed than all other methods. After followed Method 3, Method 4, Method 1 and finally Method 2 in which two (human) participants were involved, thus taking longer.

EX2 – RQ3:

Is there, among the five considered methods, one that can provide higher-overall quality ILS, taking into account both accuracy and delay as the main factors for evaluation?

By combining linguistic accuracy and delay results for each method, a consistent inverse correlation can be observed as pointed out in the EX2 discussion in Chapter 5: the least accurate method (Method 5) is the fastest one in delivering the subtitles, while the most accurate (Method 2) takes longer, up to 11 seconds of delay. The 'higher overall quality' that RQ3 sought to find is determined by the setting and the situation in which the ILS is delivered: the need for faster subtitles is to the detriment of accurate transcription and translation, or better accuracy, but slower. In terms of overall quality, EX2 shows that methods 1, 3, and 4 are a good compromise, provided accuracy can be improved when professionals are called to perform, and not students (which is indeed the case as demonstrated by the parallel research by the Universities of Vigo and Surrey). These three methods range between 6 to 7 seconds of delay, which is absolutely acceptable. In the trade-off between accuracy and delay, despite being slower, Method 2 does provide a good level of accuracy, which is hoped can be the decisive factor in choosing for the sake of correctness, rather than speed.

EX2 – RQ4:

Are there other variables to consider in searching for higher-quality methods for ILS, and what are they?

Other variables to consider in searching for a higher quality ILS method emerged from the second experiment, such as the type of setting and the cost. The industry needs to weigh up which level of accuracy is most adequate depending on the requirements of the settings: Method 3 features too many errors in recognizing and placing punctuation marks, making it difficult for audiences to read the subtitles. The more and fully-automated Methods 4 and 5 leave the final decision on what and how to broadcast the content to the machine, which can be very risky in formal and important situations, especially concerning mistranslations. It is true, though, that these are also the cheapest methods, which can be considered while evaluating which mode is the preferred.

2.3 Main RQ

Now that all four sub-RQs for each experiment are answered, the main RQ for this doctoral research can be answered too: how do different ILS methods compare in terms of quality, when quality is referred to as the joint result of both linguistic accuracy and delay in broadcasting, for English and Spanish to Italian language combinations?

The different ILS methods that were tested can be compared through contemplation of different factors. The RQ accentuates 'how' the workflows compare, rather than which is necessarily better. One of the main conclusions that may be drawn from this research is precisely that it is not only a matter of which method performs better (in accuracy, delay, or cost) but rather which is more adequate depending on the setting and the situation. All five methods tested, that were explored in all their variables, have advantages and disadvantages, present strengths and weaknesses, and it is the clients' necessities that can tip the balance in choosing one method over another. To date, however, it can be assumed that the fully-automated workflow does not meet the basic requirements of linguistic accuracy that are needed to provide a quality service in interlingual live subtitling, thus it is not advisable to use it. Nevertheless, both ASR and machine translation systems are likely to improve very much in the near future, which will possibly require testing them again to monitor their progress.

2.4 RQ5 and RQ6

Two answers remain pending regarding the last two research questions RQ5 and RQ6, which can hopefully be considered as an input in contributing to the field of respeaking and Media Accessibility training. These last questions were obtained through qualitative data collected through the preliminary and the satisfaction questionnaires for both workshops. The sample of students' replies that were considered was very much reduced and answers to both RQs are not intended to be exhaustive or comprehensive.

They only aim to detail the trainees' opinions, as they were the direct actors in the training, and provide relevant suggestions.

RQ5: Is a master's degree in Translation and Interpreting an adequate environment to train students in intra and interlingual respeaking?

Levels of performance in both experiments by the students show that they are not proficient enough to perform in interlingual respeaking. However, given that some of the more-suited profiles for respeaking are both simultaneous interpreters and subtitlers, it would indeed make sense to invest in Translation and Interpreting students to train new professionals in the field of live subtitling. A Master's degree in this area already trains several useful skills such as linguistic competence, short-term memory, rephrasing and reformulation, strategic condensation, prosody and tone, and multitasking. To exploit students' skills to train them in respeaking, be it intra or interlingual, seems to be an important step forward towards a more inclusive and accessible awareness and world. The professional future of respeaking seems promising, and is demonstrated by the potentialities shown in some of the methods tested in this research, if students were to train in some of the more specific competences required by respeaking itself. Method 3, for instance, demonstrated good potential, especially if the orality features of interpreters such as corrections, mumblings, false starts, etc. could be reduced from their output. Not only would this create a better and more recognizable audio input for ASR systems, but would also benefit the interpretation product itself, making it more pleasant (and of higher quality) to the audible audiences that listen. The integration of respeaking as part of the simultaneous interpreting academic training, therefore, does seem to be an adequate environment for such an important and necessary technique to thrive. Answers to the questionnaires demonstrate particular enthusiasm by all the students that were interested in the subject matter, and this can be viewed as a call for accessibility in a field that already embraces it, and that could with some effort achieve much more.

RQ6: Can respeaking training be useful in better educating students in Media Accessibility and, specifically, in DHOH accessibility needs?

In conjunction with the previous RQ, this last question could be considered naïve or even rhetorical, since training people (students or professionals) in respeaking would naturally educate them more on MA and DHOH accessibility. Although practice videos did not often focus on sound features that needed to be expressed as required in SDH, the topic of accessibility for non-hearing audiences was covered in depth during training. It therefore seems interesting to underline to what extent this type of training could raise awareness in this field. With the only aim of informing the readers of this thesis on the students' opinions concerning their didactic experience, it was striking how little they knew about either SDH and MA before training. When asked how much they knew about live subtitling and respeaking, or audio description for example, these seemed to be unknown fields for them. In collecting their answers with the pre-workshop questionnaire, as shown at Section 4.2, Chapter 3, almost none knew what real-time subtitling for accessibility of non-hearing people entailed, still they sought more knowledge on it. On completion of training, they were more acquainted with the topic, but they were at the same time eager to continue some form of training in it and even expressed that more features for DHOH in the subtitles would have been useful in their training. As suggested by the answer to RQ5, since it could be very effective in training translators and interpreters in this field, it is at the same time hoped that it could be fertile ground to raise awareness of the topic and train some much-needed professionals and experts in this field (Greco, 2019b).

3. Limitations, bias, and room for improvement

There were several limitations to this doctoral research, as the experiments that were conducted presented some differences in their methodologies, thus were only partially comparable for the conclusions that could be drawn from the two. In addition, especially for EX1 technical issues were encountered that invalidated some data and possibly

influenced some of the results. However, some issues could come across also in professional settings, in particular in remote ones, and this is also something that can be considered when selecting one method or the other (e.g. Method 2 carried out remotely was logistically more complex than other methods).

Regarding testing, EX1 and EX2 were originally meant to be carried out face-to-face but were remote due to the Covid-19 pandemic. This led to some important technical and organizational hindrances concerning the tasks to perform, especially for Method 2 in which two participants had to connect remotely, and because Dragon and other software had to be set up on the students' personal computers. This led to a varied range of machine performance, further to this the stability of Wi-Fi connections and audio input-output quality also had an impact, something which could have been easily avoided by installing licenses on the University's PCs and working in the classroom environment.

Although unavoidable, the students that subscribed to the workshops for both years had different levels of experience and performance in simultaneous interpreting and, consequently, in respeaking as well. This could have influenced the accuracy (and delay) results for some teams, for example, and also levels of anxiety while performing.
Lastly, subjectivity in the NTR analysis played a major role in the study. This was probably the most critical aspect the researcher dealt with during the investigation, which is also very interesting given the topic of assessment in translated and interpreted (or respoken) products. There was also a possibility that the researcher could unintentionally influence the results by carrying out analyses in a way that would have produced the expected results. It is for this reason, of course, that analyses were revised by two reviewers and, for EX2, were carried out by the participants, revised by classmates as a peer-review exercise, and finally corrected by the tutor herself. The peer-review phase was deemed particularly useful by students since it was carried out in a plenary session, not separately, therefore all could listen to the other classmates' opinions and impressions on their performance.

4. Further research developments

Some important results were achieved through cited research on feasibility of interlingual respeaking, professional profile, training model, competences and skills required, and given the novelty of the field of speech-to-text interpreting through respeaking, it is necessary that research in this field continues and thrives in the near future. Much has been said and debated on the matter of quality and assessment in translation, interpreting, and now live subtitling, and much more still remains to be investigated. Of particular interest to this research is the assessment of respoken and interpreted outputs, where subjectivity plays a major role depending on one's communicative intention and perspective.

As human-aided and fully-automated systems pave the way, it is all the more important to update research on quality outcomes deriving from different workflows, including humans and machines, to pinpoint their potential and weaknesses to better inform the industry in making a professional, well-informed choice. This is even more interesting for the sake of research given the continuous improvement in ASR and MT systems, which are likely to perform much better in the future.

Up to four language pairs have been tested in the different experiments, and it would be useful to test the different modes with highly specialized source texts, not only generic ones, to see if and how performances per method change. A major point of interest, which is essential in our opinion to make research of this kind more relevant and usable, is involving the end-users of the produced subtitles, i.e. real audiences of non-hearing or HoH subjects, and ST non-native speakers. Research on how such audiences perceive the different outputs in real time is what can make these preliminary investigations truthful.
Lastly, given the great fortune of English as a *lingua franca* at an international level, it could be of interest to test different English accents, specifically clearly non-native English speakers and see if machines succeed in processing such inputs in the same way as human respeakers and interpreters.

5. Final thoughts

To conclude, it is hoped that the results extracted through this study were relevant and meaningful for the purpose of the research itself, which was to better inform the industry on the different possibilities of creating live subtitles in varying languages, raising awareness on which issues each mode brings with it. We live in an ever-connected world in which cutting-edge technology facilitates our tasks to a great extent, in this case automated systems that improve very quickly, sometimes more than we expect. At the same time, we also think that it is crucial to have the right means to correctly interact with machines in simplifying and optimizing our work, so that an intertwining can be created between human and machine contributions without forgetting the natural endowments human perception can provide.

Bibliography

Aliprandi, C. & Verruso, F. (2006) 'HLT and real time multilingual subtitling: some state of the art in Italy and experience at the last Intersteno speech recognition championships'. Available at: https://www.intersteno.it/materiale/Forli2006/AliprandiVerrusoForli2006.pdf. (Last accessed: 12 December 2021).

Aliprandi, C., Scudellari, C., Gallucci, I., Piccinini, N., Raffaelli, M., del Pozo, A., Álvarez, A., Arzelus H., Cassaca, R., Luis, T., Neto, J., Mendes, C., Paulo, S. & Viveiros, M. (2014). 'Automatic Live Subtitling: state of the art, expectations and current trends'. Available at: https://www.vicomtech.org/upload/download/publicaciones/e3282f695262b10ecbdb197c7eb 3e319.pdf. (Last accessed: 6 November 2021).

Alonso-Bacigalupe, L. & Romero-Fresco, P. (forthcoming) 'The Role of Machine Translation in Interlingual Live Subtitling: An Empirical Qualitative Analysis'.

Arumí Ribas, M. & Romero-Fresco, P. (2008) 'A practical proposal for the Training of Respeakers'. The Journal of Specialised Translation 10, pp. 106-127. Available at http://www.jostrans.org/archive.php?display=10. (Last accessed 20 November 2021).

AVTE (2021) 'Machine Translation Manifesto'. Available at: http://avteurope.eu/wp-content/uploads/2021/09/Machine-Translation-Manifesto_ENG.pdf. (Last accessed: 13 February 2022).

Baños-Piñero, R. & Díaz-Cintas, J. (2015) 'Audiovisual translation in a global context: Mapping an ever-changing Landscape'. Hampshire: Palgrave Macmillan.

Baños- Piñero, R. (2017) 'Audiovisual translation'. In Bedijs, K. & Maaß, C. (Eds.), *Manual of romance languages in the media*. Berlin, Germany: De Gruyter, pp. 471-488.

Barik, H. (1975) 'Simultaneous interpretation: qualitative and linguistic data'. The Ontario Institute for Studies in Education, Toronto, pp. 272-297. Retrieved from: las.sagepub.com (Last accessed: 7 October 2021).

Chmiel, A., Szarkowska, A., Koržinek, D., Lijewska, A., Dutka, Ł., Brocki, Ł. & Marasek, K. (2017) 'Ear-voice span and pauses in intra- and interlingual respeaking: An exploratory study into temporary aspects of the respeaking process'. Applied Psycholinguistics, 38(5), pp. 1201-1227. Available at: https://www.cambridge.org/core/journals/applied-psycholinguistics/article/earvoice-span-and-pauses-in-intra-and-interlingual-respeaking-an-exploratory-study-into-temporal-aspects-of-the-respeaking-process/09B0B4801D381D041E5980C3E7003A6D. (Last accessed: 21 December 2021).

Davitti, E. & Sandrelli, A. (2020) 'Embracing the Complexity: A Pilot Study on Interlingual Respeaking'. *Journal of Audiovisual Translation, 3*(2), pp. 103-139.

Dawson, H. (2019) 'Feasibility, quality and assessment of interlingual live subtitling: A pilot study'. Journal of Audiovisual Translation, 2(2), pp. 36–56. Available at: https://www.jatjournal.org/index.php/jat/article/view/72. (Last accessed: 5 January 2022).

Dawson, H. (2020) 'Interlingual live subtitling – A research-informed training model for interlingual respeakers to improve access for a wide audience'. Doctoral thesis, Univerity of Roehampton, London.

Dawson, H. (2021) 'Exploring the quality of different live subtitling methods: a Spanish to English follow up case study'. Paper presented on 17th September 2021 at the 7th IATIS conference, Pompeu Fabra University (Barcelona, Spain).

Dawson, H., & Romero-Fresco, P. (2021) 'Towards research-informed training in interlingual respeaking: an empirical approach'. The Interpreter and Translator Trainer, 15(1), pp. 66-84. Available at: https://www.tandfonline.com/doi/full/10.1080/1750399X.2021.1880261. (Last accessed: January 2022).

De Bot, K. (2000) 'Language proficiency effects in translation and SI'. In Englund Dimitrova, B., & Hyltenstam, K. (Ed), *Language processing and SI, interdisciplinary perspectives*. Amsterdam & Philadelpia, John Benjamins, pp. 80-88.

Díaz-Cintas, J. (2005) 'Audiovisual translation today. A question of accessibility for all'. *Translating Today Magazine*, 4, pp. 3-5.

Díaz-Cintas, J. & Remael, A. (2007) 'Audiovisual Translation: Subtitling'. St Jerome Publishing, Manchester.

Díaz-Cintas, J. (2009) 'New Trends in Audiovisual Translation'. (Ed), Topics in Translation, Bristol, Buffalo, Toronto.

Díaz-Cintas, J. & Remael, A. (2021) 'Subtitling: Concepts and Practices'. Routledge, London & New York.

Dumouchel, P., Boulianne, G., & Brousseau, J. (2011) 'Measures for quality of closed captioning' in Şerban, A., Matamala, A. & Lavaur, J.M. (Eds.) *Audiovisual Translation in Close-up: Practical and Theoretical Approaches*. Bern: Peter Lang, pp. 161-172.

Eichmeyer-Hell, D. (2021) 'Speech recognition (Respeaking) vs. the conventional method (Keyboard): A quality-oriented comparison of speech-to-text interpreting techniques and addressee preferences'. Paper presented on 17th September 2021 at the 7th IATIS conference, Pompeu Fabra University (Barcelona, Spain).

Eichmeyer-Hell, D. (forthcoming) 'WIRA: Model for qualitative assessment of a speech-to-text interpreting service, taking into account the user perspective'. PhD Thesis, University of Graz.

Eppler, E. D., & Ozón, G. (2013). 'English words and sentences: An introduction'. Cambridge: Cambridge University Press.

Errico, E. (2016) 'Qualità in interpretazione, modalità interpretative e ruolo dell'interprete: alcune osservazioni'. In *Estudios de Traducción* 6, Ediciones Complutense, pp. 163-178. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiVmYaWmLX3AhUt_rsIHc_NCOkQFnoECAkQAQ&url=https%3A%2F%2Frevistas.ucm.es%2Findex.php%2FESTR%2Farticle%2Fdownload%2F53010%2F48663&usg=AOvVaw3syLUEW7Gsn7esiXzD67OH. (Last accessed: 5 December 2021).

Errico, E. & Morelli, M. (2015) 'Le sfide della qualità in interpretazione. Ricerca, didattica e pratica professionale'. Lingua, traduzione e didattica, Franco Angeli, Milano.

Errico, E., Anfuso, M., Pagano, A. & Pongiglione, B. (2021) 'Il *role-play* nella formazione in interpretazione dialogica: alcune osservazioni sull'omissione' (online). In Trovato, G. (ed), Quaderno 17, AGON, Rivista Internazionale di Studi Culturali, Linguistici e Letterari. Available at: http://agon.unime.it/files/2021/06/S1704.pdf. (Last accessed: 13 July 2021).

Eugeni, C. (2006a) 'Introduzione al rispeakeraggio televisivo'. In Eugeni, C. & Mack, G. (ed) *Proceedings of the first international seminar on real-time intralingual subtitling.* inTRAlinea, Special issue. Available at: https://www.intralinea.org/specials/article/1683. (Last accessed: 10 December 2021).

Eugeni, C. & Mack, G. (2006) 'Proceedings of the first international seminar on real-time intralingual subtitling'. inTRAlinea, Special issue. Available at: http://www.intralinea.org/specials/article/Proceedings_of_the_First_International_Seminar_on_real-time_intralingual_su. (Last accessed: 6 January 2022).

Eugeni, C. (2007) 'Il rispeakeraggio televisivo per sordi: per una sottotitolazione mirata del TG'. *InTRAlinea* 8. Available at: http://www.intralinea.org/archive/article/Il_rispeakeraggio_televisivo_per_sordi (Last accessed: 15 September 2021).

Eugeni, C. (2008) 'La sottotitolazione in diretta TV; Analisi strategica del respeakeraggio verbatim di BBC News'. Doctoral Thesis, Università degli Studi di Napoli Federico II, Italy.

Eugeni, C. (2012) 'Measuring Audiovisual Translation. A model for the analysis of intralingual live subtitles'. US- China Foreign Language, 10, pp. 1276-1286.

Eugeni, C. (2017) 'La sottotitolazione intralinguistica automatica'. CoMe 2(1), Studi di Comunicazione e Mediazione linguistica e culturale. *II*(1), pp. 102-113.

Eugeni, C. (2019) 'Technology in court reporting – Capitalising on human-computer interaction''. In International Justice Congress Proceedings, Uluslararası Adalet Kongresi (UAK), 2-4 May 2019, pp. 873-881.

Eugeni, C., & Bernabé, R. (2019) 'LTA project – Bridging the gap between training and the profession in real-time intralingual subtitling'. In Linguistica Antverpiensia New Series*: Themes in Translation Studies*, 18. Available at: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/20. (Last accessed: September 2022).

Eugeni, C. (2020) 'Human-Computer Interaction in Diamesic Translation. Multilingual Live Subtitling'. In: Dejica, D., Eugeni, C. & Dejica-Cartiș, A., (eds.) *Translation Studies and Information Technology – New Pathways for Researchers, Teachers and Professionals*. Translation Studies Series. Editura Politehnica, Politehnica University Timișoara, pp. 19-31.

Eugeni, C. (2021) 'Live editing in live parliamentary subtitling – Repercussions on accuracy and delay'. Presentation for the International Conference Media for All, *Sketching Tomorrow's Mediascape* (online). Barcelona, 27-29 January 2021. Available at:

https://www.youtube.com/watch?v=gnER6OO-pPg. (Last accessed: 17 May 2021).

European Broadcasting Union (EBU) (2004) '*EBU Report Access Services*, EBU Technical Information.

Fantinuoli, C., & Prandi, B. (2021) 'Towards the evaluation of automatic simultaneous speech translation from a communicative perspective'. *Proceedings of the 18th International Conference on Spoken Language Translation* (online). Bangkok, Thailand, August 5-6, 2021. Association for Computational Linguistics, pp. 245-254.

Freddi, M. & Luraghi, S. (2011) 'Titling for the Opera House: A Test Case for Universals of Translation?'. In Incalcaterra McLoughlin, L., Biscio, M. & Ní Mhainnín, M.A. (eds), *Audiovisual Translation Subtitles and Subtitling – Theory and Practice*. New Trends in Translation Studies, 9, Peter Lang, pp. 55-85.

Fryer, L. (2016) 'An Introduction to Audio Description – A practical guide'. London, Routledge.

Gambier, Y., Gile, D. & Taylor, Christopher (1997) (eds) 'Conference Interpreting: Current Trends in Research'. Amsterdam & Philadelphia: John Benjamins.

Gambier, Y. & Gottlieb, H. (2001) (eds), '(Multi)Media Translation: Concepts, Practices and Research'. Amsterdam: John Benjamin, pp. 259-74.

Georgakopoulou, P. (2013) 'Il progetto europeo SUMAT – Sottotitolazione assistita dalla traduzione automatica'. In Eugeni, C. & Zambelli, L. (eds.) *Respeaking*. Specializzazione online, pp. 120-122. Available at: https://www.accademia-aliprandi.it/public/specializzazione/respeaking.pdf. (Last accessed: 16 January 2022).

Georgakopoulou, P. (2019) 'Technologization of Audiovisual Translation'. In Perez-González (Ed.), *The Routledge Handbook of Audiovisual Translation*. London & New York: Routledge, pp. 516-539.

Gerver, D. (1969/2002). 'The effects of source language presentation rate on the performance of simultaneous conference interpreters'. In Pöchhacker, F. & Shlesinger, M. (Eds), *The Interpreter Studies reader*. London: Routledge, pp. 53-66.

Gile, D. (1992) 'Basic Theoretical Components in Interpreter and Translator Training'. In Dollerup, C. & Loddegaard, A. (Eds). *Teaching Translation and Interpreting: Training, Talent and Experience*. Amsterdam & Philadelphia: John Benjamins.

Gile, D. (1995) 'Basic Concepts and Models for Interpreter and Translator Training'. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Gile, D. (2005) 'Directionality in Conference Interpreting: a cognitive view'. In Godijns, R. & Hindedael, M. (eds), *Directionality in Interpreting. The 'Retour' or the Native?*, Ghent: Communication and Cognition, pp. 9-26. Available at: https://www.researchgate.net/publication/260120149_Directionality_in_conference_interpreting_a_cognitive_view. (Last accessed: 20 January 2022).

Gile, D. (2009) 'Basic concepts and models for interpreter and translator training'. Amsterdam: John Benjamins.

Gile, D. (2011) 'Errors, omissions and infelicities in broadcast interpreting: preliminary findings from a case study'. In Alvstad, C., Hild, A. & Tiselius, E. (eds), *Methods and strategies of process research: integrative approaches in translation studies*. Amsterdam & Philadelphia: John Benjamins, pp. 201-218.

Gile, D. (2015) 'Effort models'. In Pöchhacker, F. (Ed.), Routledge encyclopedia of interpreting studies. London: Routledge, pp. 135-137.

Greco, G.M. (2016) 'On accessibility as a human right, with an application to media accessibility'. In Matamala, A. & Orero, P. (eds.). *Researching Audio Description: New Approaches*. London: Palgrave Macmillan, pp. 11-33.

Greco, G.M. (2018). 'The nature of accessibility studies'. *Journal of Audiovisual Translation, 1*(1), pp. 205-232.

Greco, G.M. (2019a) 'Accessibility Studies: Abuses, misuses and the method of poetic design'. In Stephanidis, C. (ed.) *HCI International 2019 – Late Breaking Papers*. Springer International Publishing: New York, pp. 15-27.

Greco, G.M. (2019b). 'Towards a pedagogy of accessibility: The need for critical learning spaces in media accessibility education and training'. Linguistica Antverpiensia, New Series: *Themes in Translation Studies*, *18*, pp. 23-46.

Greco, G.M. & Jankowska, A. (2020) 'Media Accessibility Within and Beyond Audiovisual Translation'. In Bogucki, Ł., Deckert, M. (eds.), *The Palgrave Handbook of Audiovisual Translation and Media Accessibility*, Palgrave Studies in Translating and Interpreting, pp. 57-81. Available at: https://doi.org/10.1007/978-3-030-42105-2_4. (Last accessed: June 2022).

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. & Verplank, W. (1992) 'ACM SIGCHI curricula for human-computer interaction. Report of the ACM SIGCHI Curriculum Development Group'. Broadway, ACM. Available at: https://www.researchgate.net/publication/234823126_ACM_SIGCHI_curricula_for_human-computer_interaction. (Last accessed: 12 July 2021).

Hurtado Albir, A. (2017) 'Researching translation competence by PACTE Group'. Amsterdam & Philadelphia: John Benjamins, pp. 187-189.

Jakobson, R. (1959) 'On linguistic aspects of translation'. In Venuti, L. (Ed) 2000. *The translation Studies Reader*, London & New York: Routledge, pp. 113-118.

Kalina, S. (2015) 'Compression'. In Pöchhacker, F. (Ed.), *Routledge encyclopedia of interpreting studies*, London: Routledge, pp. 73-75.

Koponen, M. & Salmi, L. (2015) 'On the Correctness of Machine Translation: A Machine Translation Post-Editing Task'. *The Journal of Specialised Translation, 23*, pp. 118-136. Available at: https://jostrans.org/issue23/art_koponen.pdf. (Last accessed 13 February 2022).

Kurz, I. (1992) 'Shadowing' Exercises in Interpreting Training'. In Dollerup, C. & Loddegaard, A. (eds), *Teaching Translation and Interpreting. Training, Talent and Experience*, Amsterdam & Philadelphia: John Benjamins, pp. 245-250.

Lambert S. (1989) 'La formation d'interprètes: la méthode cognitive'. Meta 34 (4), pp. 736-744.

Lambert S. (1992) 'Shadowing'. The Interpreters' Newsletter 4, pp. 15-24.

Lambourne, A., Hewitt, J. & Warren, S. (2004) 'Speech-Based Real-Time Subtitling Services'. International Journal of Speech Technology 7, Kluwer Academic Publishers, The Netherlands, pp. 269-279.

Lambourne, A. (2006) 'Subtitle respeaking'. In Eugeni, C & Mack, G. (eds.) *inTRAlinea, Special Issue on Respeaking*. Available at: http://www.intralinea.org/specials/article/Subtitle_respeaking. (Last accessed: 17 December 2021).

Lambourne, A. (2007) 'Real-Time subtitling: Extreme Audiovisual Translation'. LSP Translation Scenarios, Vienna.

Leonardi, V. (2000) 'Equivalence in Translation: Between Myth and Reality'. Translation Journal online, 4(4). Available at: https://translationjournal.net/journal/14equiv.htm. (Last accessed: September 2021).

Mack, G. (2006a) 'Detto scritto: un fenomeno, tanti nomi'. In Eugeni, C. & Mack, G. (ed) *Proceedings of the first international seminar on real-time intralingual subtitling*. inTRAlinea, Special issue. Available at: https://www.intralinea.org/specials/article/Detto_scritto_un_fenomeno_tanti_nomi. (Last accessed: 10 December 2021).

Manetti, I. (2018) 'L'interaction homme-machine – analyse d'un cas de sous-titrage interlinguistique semi-automatisé'. CoMe, 3(1), pp. 65-80. Available at:

http://comejournal.com/wp-_content/uploads/2019/03/5.-I.-Manetti-CoMe-III-1-1.pdf. (Last accessed: 20 November 2021).

Matamala, A., Oliver, A., Álvarez, A., Azpeitia, A. (2015) 'The reception of intralingual and interlingual automatic subtitling: an exploratory study within the HBB4ALL project'. In *Proceedings of the 37th Conference Translating and the Computer*, 12-17, London, 26-27th November.

Matamala, A., Romero-Fresco, P. & Daniluk, L. (2017) 'The Use of Respeaking for the Transcription of Non-Fictional Genres – An Exploratory Study'. inTRAlinea, 19. Available at: http://www.intralinea.org/archive/article/the_use_of_respeaking_for_the_transcription_of_non_fictional_genres. (Last accessed: 26 November 2021).

McIntyre, D., Moores, Z. & Price, H. (2018). 'Respeaking Parliament: Using Insights from Linguistics to Improve the Speed and Quality of Live Parliamentary Subtitles'. Huddersfield: Language Unlocked/Institute for Applied Linguistics.

MIUR (2012). 'Requisiti per il riconoscimento della validità delle certificazioni delle competenze linguistico-comunicative in lingua straniera del personale scolastico'. Available at: https://www.miur.gov.it/documents/20182/0/Decreto+Ministeriale+7+marzo+2012/80d5df76-5247-49f1-9785-b81d74b4a4f8?version=1.0. (Last accessed: 23 September 2021).

Moores, Z. (2020) 'Fostering access for all through respeaking at live events'. JoSTrans, Journal of Specialised Translation, 33. Available at: https://jostrans.org/issue33/art_moores.php. (Last accessed: 5 November 2021).

Neves, J. (2005) 'Audiovisual Translation: Subtitling for the deaf and hard-of-hearing'. Doctoral thesis, Roehampton University, United Kingdom.

Nida, E. A. (1964/2003) 'Toward a Science of Translating'. Leiden: Brill Archive.

Nida, E. A. & Taber, C.R. (1969/1982) 'The Theory and Practice of Translation'. Leiden: Brill Archive.

Ofcom. (2005) 'Subtitling – An issue of speed?'. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0018/16119/subt.pdf. (Last accessed: 3 November 2021).

Ofcom. (2013) 'Measuring the quality of live subtitling'. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0017/51731/qos-statement.pdf. (Last accessed: September 2021).

Ofcom. (2014a) 'Measuring live subtitling quality: results from the first sampling exercise'. Available at: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-_subtitling/sampling-results. (Last accessed: November 2021).

Ofcom. (2014b) 'Measuring live subtitling quality: results from the second sampling exercise'. Available at: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-__subtitling/sampling-results-2. (Last accessed: November 2021).

Ofcom. (2015a) 'Measuring live subtitling quality: results from the third sampling exercise'. Available at: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-_subtitling/sampling-results-3. (Last accessed: November 2021).

Ofcom. (2015b) 'Measuring live subtitling quality: results from the fourth sampling exercise'. Available at: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-__subtitling/sampling_results_4. (Last accessed: November 2021).

Orero, P. (2004) 'Audiovisual Translation: A new dynamic umbrella'. In Orero, P. (ed.), *Topics in Audiovisual Translation*. Amsterdam: John Benjamins, VII.

Pagano, A. (2020a) 'Verbatim vs. Edited Live Parliamentary Subtitling' in Dejica, D., Eugeni, C., Dejica-Cartis, A. *Translation Studies and Information Technology – New Pathways for researchers, teachers and professionals*, Editura Politehnica Timisoara, pp. 32-44.

Pagano, A. (2020b) 'Terminologia e Interpretazione di tribunale: un'esperienza didattica spagnolo-italiano'. Publifarum, 33. Università di Genova. Available at: https://www.publifarum.farum.it/index.php/publifarum/article/view/435. (Last accessed: 17 August 2021).

Perego, E. & Taylor, C. (2012) 'Tradurre l'audiovisivo'. Carocci editore, Roma.

Pérez-González, L. (2008) 'Audiovisual translation'. In Baker, M. & Saldanha, G. (Eds.), *Routledge encyclopedia of translation studies*. 2nd ed. London, New & York: Routledge, pp. 13-20.

Pérez-González, L. (2019) 'Routledge Handbook of Audiovisual Translation'. (Ed), London & New York: Routledge.

Pöchhacker, F., (2004) 'Introducing Interpreting Studies'. London: Routledge.

Pöchhacker F. (2008) 'The turns of Interpreting Studies', in Hansen, G., Chesterman, A., Gerzymisch-Arbogast, H. (Eds) *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*, Amsterdam & Philadelphia, John Benjamins, pp. 25-46.

Pöchhacker, F., (2013) 'Researching Quality: a two-Pronged Approach'. In García Becerra, O., Pradas Macías, E. M. & Barranco Droege, R. (eds.), *Quality in Interpre- ting: Widening the Scope*, Vol. 1, Granada: Comares, pp. 33-56.

Pöchhacker, F. & Remael, A. (2019) 'New efforts? A competence-oriented task analysis of interlingual live subtitling'. Linguistica Antverpiensia, New Series: *Themes in Translation Studies*, 18, pp. 130-143.

Pym, A. (2008) 'On omission in simultaneous interpreting: risk analysis of a hidden effort'. In Hanses, G., Chesterman, A. & Gerzymisch-Arbogast, H., (eds), *Efforts and models in interpreting and translation research: a tribute to Daniel Gile*, Amsterdam & Philadelphia: John Benjamins, pp. 83-105.

Pym, A., (2010) 'Exploring Translation Theories', London & New York: Routledge.

RAI Italia (2021) 'Norme e convenzioni editoriali essenziali – sottotitoli televisivi per spettatori sordi e con difficoltà uditive'. Available at: https://www.rai.it/dl/doc/2020/10/19/1603121663902_PREREGISTR_22_feb_2016_-_Norme_e__Convenzioni_essenziali_per_la_composiz...%20-%20Copia.pdf. (Last accessed: 4 February 2022).

Remael, A. & van der Veer, B. (2006) 'Real-time Subtitling in Flanders: Needs and teaching', inTRAlinea, Special Issue. Available at: http://www.intralinea.org/specials/article/Real-Time_Subtitling_in_Flanders_Needs_and_Teaching. (Last accessed: 27 December 2021).

Remael, A. (2012) 'Media Accessibility'. In Gambier, Y. & van Doorslaer, L. (Eds.), *Handbook of Translation Studies*, vol. 3, pp. 95-101. Amsterdam, Philadelphia, John Benjamins Publishing Company.

Remael, A., Reviers, N., & Vandekerckhove, R. (2016). 'From Translation Studies and Audiovisual Translation to Media Accessibility'. *Target*, 28(2), pp. 248-260.

Retief, M. & Letsosa, R. (2018) 'Models of disability: A brief overview'. *HTS Teologiese Studies/ Theological Studies* 74(1). Available at: https://hts.org.za/index.php/hts/article/view/4738. (Last accessed: 15 November 2021).

Riccardi, A. (2015) 'Speech rate'. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies*. London: Routledge, pp- 397-399.

Robert, I. & Remael, A. (2017) 'Assessing quality in live interlingual subtitling: A new challenge'. Linguistica Antverpiensa, New Series: *Themes in Translation Studies, 16*, pp. 168-195. Available at: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/454. (Last accessed: 15 December 2021).

Robert, I., Schrijver, I. & Diels, E. (2019a) 'Live subtitlers who are they? A survey study', Linguistica Antverpiensia, New Series: *Themes in Translation Studies,* 18, pp. 101-129. Available at: https://lans- tts.uantwerpen.be/index.php/LANS-TTS/article/view/544. (Last accessed: 30 November 2021).

Robert, I., Schrijver, I. &Diels, E. (2019b) 'Trainers' and employers' perception of training in intralingual and interlingual live subtitling: A survey study'. Journal of Audiovisual Translation, 2(1), pp. 1-25. Available at: https://www.jatjournal.org/index.php/jat/article/view/61. (Last accessed: 20 January 2022).

Romero-Fresco, P. (2009) 'More haste less speed: Edited vs. verbatim respeaking', *International Journal of Applied Linguistics* (VIAL), pp. 109–33.

Romero-Fresco, P. (2010) 'Standing on quicksand: hearing viewers' comprehension and reading patterns of respoken subtitles for the news'. In Díaz-Cintas, J., Matamala, A. & Neves, J. (ed), *New Insights into Audiovisual Translation and Media Accessibility*, *Media for All 2*, Editions Rodopi, B. V., Amsterdam & New York, pp. 175-194.

Romero-Fresco, P. (2011) 'Subtitling through speech recognition: Respeaking'. St. Jerome Publishing, Manchester.

Romero Fresco, P. (2012) 'Respeaking in translator training curricula. Present and future prospects'. *The Interpreter and Translator Trainer*, 6(1), pp. 91-112. Available at: https://www.tandfonline.com/doi/abs/10.1080/13556509.2012.10798831?journalCode=ritt20. (Last accessed: 3 December 2021).

Romero-Fresco, P. (2015) 'Respeaking as a bridge between subtitling and interpreting'. In Pöchhacker, F. (ed.) *Routledge Encyclopaedia of Interpreting Studies*. New York & London: Routledge, pp. 349-351.

Romero-Fresco, P. & Martínez, J. (2015) 'Accuracy rate in live subtitling: The NER model'. In Díaz-Cintas, J. & Baños, R. (ed.) *Audiovisual Translation in a Global Context: Mapping an Ever-changing Landscape*. London & New York: Palgrave Macmillan, pp. 28-50.

Romero-Fresco, P. & Pöchhacker, F. (2017) 'Quality assessment in interlingual live subtitling: The NTR model'. Linguistics Antverpiensia, New Series: Themes in Translation Studies, 16, pp. 149-167. Available at: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/438. (Last accessed: 3 January 2022).

Romero-Fresco, P. (2018) 'In support of a wide notion of Media Accessibility: Access to content and access to creation'. Journal of Audiovisual Translation, 1(1), pp. 187-204. Available at: http://www.jatjournal.org/index.php/jat/article/view/53/12. (Last accessed: 16 December 2021).

Romero-Fresco, P. (2019a) 'Accessible filmmaking: translation and accessibility from production' in Perez-González, P. (ed.) *The Handbook of Audiovisual Translation*. Oxon & New York: Routledge, pp. 498– 515.

Romero-Fresco, P. (2019b) 'Respeaking: subtitling through speech recognition' in Perez-González, L. (ed.), *The Routledge Handbook of Audiovisual Translation*. Oxon & New York: Routledge, pp. 96–113.

Romero-Fresco, P. & Eugeni, C. (2020) 'Live subtitling through respeaking'. In Bogucki, Ł. & Deckert, M. (eds.), *Handbook of Audiovisual Translation and Media Accessibility*. Palgrave MacMillan, pp. 269-295.

Romero-Fresco, P. (2021a). 'Creative media accessibility: Placing the focus back on the individual'. In Antona, M. & Stephanidis, C. (Eds.), *Lecture Notes in Computer Science*, pp. 291-307.

Romero-Fresco, P. & Alonso-Bacigalupe, L. (2021) 'Testing the efficiency of different live (sub)tiling methods: a practical experience'. Paper presented on 17th September 2021 at the 7th IATIS conference, Pompeu Fabra University (Barcelona, Spain).

Romero-Fresco, P., & Alonso-Bacigalupe, L. (2022). 'An Empirical Analysis on the Efficiency of Five Interlingual Live Subtitling Workflows'. XLinguae, Special issue edited by Ł. Bogucki & M. Deckert.

Russello, C. (2010) 'Teaching respeaking to conference interpreters'. Intersteno.it (online). Education committee archive. Available at: https://www.intersteno.it/materiale/ComitScientifico/EducationCommittee/Russello2010Teaching%20Respeaking%20to%20Conference%20Interpreters.pdf. (Last accessed: 13 January 2022).

Russello, C. (2013) 'Aspetti didattici'. In Eugeni, C. & Zambelli, L. (eds.) *Respeaking*. Specializzazione online, pp. 44-53. Available at: https://www.accademia-aliprandi.it/public/specializzazione/respeaking.pdf. (Last accessed: 12 January 2022).

Sandrelli, A. (2020) 'Interlingual respeaking and simultaneous interpreting in a conference setting: a comparison'. Spinolo, N. & Amato, A. (ed) *Technology in Interpreter Education and Practice*. inTRAlinea, Special issue. Available at: https://www.intralinea.org/specials/article/2518. (Last accessed: 5 January 2022).

Seleskovitch, D., & Lederer, M. (1989) 'Pédagogie Raisonnée de l'Interprétation'. Paris: Didier Érudition. Translated by Harmer, J. (1995) as 'A Systematic Approach to Teaching Interpretation'. Silver Springs, MD: Registry of Interpreters for the Deaf Publications.

Snyder, J. (2008) 'Audio description: The visual made verbal'. In Díaz-Cintas, J. (Ed), *The Didactics of Audiovisual Translation*. Amsterdam & Philadelpia, John Benjamins, pp. 191-198.

Soria, E. (2016) 'A proposal to assess accuracy for interlingual respeaking (EN>ES). Adapting the NER Model'. MA dissertation, University of Roehampton.

Szarkowska, A., Dutka, Ł., Krejtz, K. & Pilipczuk, O. (2017) 'Respeaking crisis points. An exploratory study into critical moments in the respeaking process'. In Deckert, M. (ed.) *Audiovisual Translation – Research and Use.* Bern: Peter Lang, pp. 179-201.

Szarkowska, A., Krejtz, K., Dutka, Ł. & Pilipczuk, O. (2018) 'Are interpreters better respeakers?', *The Interpreter and Translator Trainer*, 12(2), pp. 207-226. Available at: https://www.tandfonline.com/doi/full/10.1080/1750399X.2018.1465679. (Last accessed: 27 December 2021).

Szarkowska, A. & Wasylczyk, P. (2018) 'Five things you wanted to know about Audiovisual Translation research, but were afraid to ask'. Journal of Audiovisual Translation, 1(1), pp. 8-25. Available at: http://www.jatjournal.org/index.php/jat/article/view/41. (Last accessed: 14 December 2021).

Viezzi, M. (1999) 'Aspetti della qualità dell'interpretazione'. In Falbo, C., Russo, M. & Straniero Sergio, F. (eds.), *Interpretazione simultanea e consecutiva. Proble- mi teorici e metodologie didattiche*, Milano: Hoepli, pp. 140-151.

Vinay, J.P. & Darbelnet, J. (1995) 'Comparative Stylistics of French and English: A Methodology for Translation'. Amsterdam & Philadelphia: John Benjamins.

Wasserman, D., & Aas, S., Blustein, J., Putnam, D. (2002) 'Disability: Definitions and Models'. *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (ed.). Available at: https://plato.stanford.edu/archives/sum2022/entries/disability/. (Last accessed: 3 November 2021).

Webography and Filmography

AIIC, Association Internationale des Interprètes de Conférence. Available online at: https://aiic.org. (Last accessed: 9 December 2021).

Arnáiz Uzquiza, V. (2014) 'UNE-153010:2012. Norma Española de subtitulado para sordos'. Available online at: https://uvadoc.uva.es/handle/10324/5677. (Last accessed: August 2021).

Doshi, K. (2021) 'Audio Deep Learning Made Simple: Automatic Speech Recognition (AS), How it Works'. Available online at: https://towardsdatascience.com/audio-deep-learning-made-simple-automatic-speech-recognition-asr-how-it-works-716cfce4c706. (Last accessed: 15 June 2022).

DTV4ALL Project, Digital Television for All (2007-2013). Available at: https://cordis.europa.eu/project/id/224994/it. (Last accessed: 26 November 2021).

Encyclopedia Britannica online. Available online at http://www.britannica.com/. (Last accessed: 12 December 2021).

European Commission website, Definition of Departments and executive agencies. 'Conference interpreting – types and terminology'. Available online at: https://ec.europa.eu/info/departments/interpretation/conference-interpreting-types-and-terminology_en. (Last accessed: 20 July 2021).

European Commission Speech Repository. (2008) 'El reciclaje de medicamentos'. Available online at: https://webgate.ec.europa.eu/sr/speech/el-reciclaje-de-medicamentos. (Last accessed: November 2021).

European Commission_Speech Repository. (2008) 'Marco jurídico de las comunicaciones electrónicas'. Available online at: https://webgate.ec.europa.eu/sr/speech/marco-jur%C3%ADdico-de-las-comunicaciones-electrónicas. (Last accessed: December 2021).

European Commission Speech Repository. (2009) 'Los peligros del teléfono móvil'. Available online at: https://webgate.ec.europa.eu/sr/speech/los-peligros-del-teléfono-móvil. (Last accessed: November 2021).

European Commission Speech Repository. (2009) 'Mobile phones'. Available online at: https://webgate.ec.europa.eu/sr/speech/mobile-phones. (Last accessed December 2020).

European Commission Speech Repository. (2019) 'The future of work'. Available online at: https://webgate.ec.europa.eu/sr/speech/future-work. (Last accessed December 2020).

European Commission Speech Repository. (2019) 'Friday 13th'. Available online at: https://webgate.ec.europa.eu/sr/speech/friday-13th. (Last accessed December 2020).

ILSA Project, Interlingual Live Subtitling for Access (2017-2021). Available online at: http://galmaobservatory.eu/projects/interlingual-live-subtitling-for-access-ilsa. (Last accessed: November 2021).

ILSA, Interlingual Live Subtitling for Access (2019) 'Respeaking in Italy – An interview with Carlo Eugeni'. Available online at: https://www.youtube.com/watch?v=XXsoo5xhS-M. (Last accessed: August 2021).

LTA Project, Live Text Access (2021). Available online at: https://ltaproject.eu/. (Last accessed: 8 December 2021).

SMART, Shaping Multilingual Access through Respeaking Technology (2017-2019). Available online at: http://galmaobservatory.eu/projects/shaping-multilingual-access-through-respeaking-technology-smart/. (Last accessed: December 2021).

Speech by Antonio Banderas en la ceremonia de los Premios Goyas. Available online at: https://www.youtube.com/watch?v=ZqJOr8Ba90U. (Last accessed: December 2022).

Speech by King Felipe VI (2014). Available online at: https://www.youtube.com/watch?v=Dk4ZfvoAWNs. (Last accessed: December 2022).

Speech by Pope Francisco. Available online at: https://www.youtube.com/watch?v=LiTJvHmFtbE. (Last accessed: February 2022).

Speech by President José Mujica, Río +20. Available online at: https://www.youtube.com/watch?v=3cQgONgTupo. (Last accessed: December 2022).

Speech by President Mariano Rajoy. Available online at: https://www.youtube.com/watch?v=iSzk2Sm4Fl4. (Last accessed: December 2022).

Speech by Italian Senator Liliana Segre at the European Parliament (2020). Available online at: https://www.youtube.com/watch?v=o56HpNydkog. (Last accessed: 15 January 2022).

Speech by Italian actor Roberto Benigni, retrieved from 'I dieci comandamenti'. Available online at: https://www.youtube.com/watch?v=u9x2aXyBVDI. (Last accessed 17 December 2021).

TEDx Talk Ocala. (2015) 'The secret of becoming mentally strong'. Available online at: https://www.youtube.com/watch?v=TFbv757kup4. (Last accessed December 2020).

TEDx Talk Stockholm (2018) – Greta Thunberg 'School strike for climate – save the world by changing the rules'. Available online at: https://www.ted.com/talks/greta_thunberg_school_strike_for_climate_save_the_world_by_changing_the_rules/transcript?language=it. (Last accessed August 2021).

TEDx Talks Gijon. (2018) 'Lo positivo de fracasar en el amor'. Available online at: https://www.youtube.com/watch?v=xbX3NOiRX_w&t=2s. (Last accessed: December 2021).

TEDx Talks Youth Madrid. (2015) 'La pasión de ser diferente'. Available online at: https://www.youtube.com/watch?v=rJY_8oN6gIs. (Last accessed: January 2022).

TEDx Talks Padova. 'Meditazione 3.0, la cura del pensiero'. Available online at: https://www.youtube.com/watch?v=MmTWTGHRwfk&t=1s. (Last accessed 10 January 2022).

TEDx Talks Rimini. 'Che cosa vuol dire "vero" su Instagram?'. Available online at: https://www.youtube.com/watch?v=fDSeiL1RgGM. (Last accessed 25 January 2022).

TEDx Talks Vicenza. 'Chi è Salvatore Aranzulla'. Available online at: https://www.youtube.com/watch?v=QU1L-AHHEh0. (Last accessed: 20 January 2022).

NDACA (National Disability Arts Collection and Archive) (2019). 'Social Model of Disability'. Available online at: https://www.youtube.com/watch?v=24KE__OCKMw. (Last accessed: 2 November 2021).

Appendix 1

First experiment materials
(NTR analyses, video and audio recordings)

All analyses carried out for the five teams in the first experiment (EX1, English to Italian) are available at the following cloud link: https://drive.google.com/drive/folders/1jYZy5RvfFyV5LFqjk-uw4FZkrvrfMWkU?usp=sharing.

Folders are divided into teams from 1 to 5, and each Team is divided into Method 1, Method 2, and Method 3. For each method, audiovisual material recorded during the experiment is available: .mp3 recording, .mp4 videos (both source text and participant feeds, and with separated audio tracks). For each method also the NTR analysis template, with the transcription file is available in the dedicated folder 'Analysis'.

Appendix 2

Pre-workshop questionnaires for the first experiment

Pre-workshop explorative questionnaires

The explorative questionnaire below is anonymous and is designed to collect some non-sensitive information about the participants for the purposes of the project.

General information and academic background

Sex: Male Female Other

Age: 20 21 22 23 24 25 26 27 28 Other

Besides English, what is your major language in your academic curriculum?

Did you attend your bachelor's degree course at the University of Genoa? Yes No

If you marked "No" to the previous question, please indicate from which University you graduated for your bachelor's degree.

Regardless of the University, which Italian bachelor's degree did you graduate from?

- Lingue e culture moderne o ex Lingue e letterature straniere (L-11)
- Teorie e tecniche della mediazione interlinguistica (L-12)
- Scienze politiche e delle relazioni internazionali (L-36)
- Lettere (L-10)
- Scienze del turismo (L-15)
- Other

If you marked "Other" to the previous question, indicate which bachelor's degree you graduated from, specifying the class (L-...).

In your bachelor's or master's degree program, have you taken formal courses, modules, or workshops in interlingual or interlingual subtitling for the deaf, before this respeaking workshop?
('Formal' is meant by courses with a minimum number of hours with compulsory attendance and activities, with a certification, a final exam or test, and so on).

If you have answered positively to the previous question, please specify approximately the number of hours, ECTS, certifications foreseen, or the activities you have carried out, detailing for interlingual subtitling which language pair you worked with.

Have you taken any courses or lectures in interpreting and, specifically, simultaneous interpreting prior to the respeaking workshop?
(Here we refer to courses in any language, not just English).

In your academic record, could you roughly quantify the total number of hours of simultaneous interpreting practice before this respeaking course? You can also specify whether they were semester-long or year-long courses, how many hours per week, and so on.
(Here we refer to the number of hours of presence or distance learning, not to autonomous practice, and of any language, not just English).

Making now reference to the English language only, could you quantify approximately in your academic career the number of hours of courses dedicated to simultaneous interpreting before the respeaking course?
(Here we refer to the number of hours of presence or distance learning, not to autonomous practice).

Before the course, did you ever hear of respeaking?

- Yes, I perfectly knew what it was
- Yes, but I did not know exactly what it was
- Never

If yes, how did you come to know about the technique of respeaking?
(More than one option can be selected).

- From university professors or professionals
- In courses, modules, workshops or exams at my University
- Through job offers
- From classmates or colleagues
- From online resources (articles, scientific publications, research, etc.)
- Other

In your bachelor's or master's degree program, have you ever attended a module/workshop or other related to respeaking prior to this training? Yes No

Why did you sign up for this workshop of respeaking?
(More than one option can be selected).

- I did not know what it was and I wanted to find out
- I heard of it and I wanted to know more about it
- I knew what it was and I wanted to subscribe for a specific training in it
- I knew what it was and I wanted to know more about it for future job opportunities

- I needed the ECTS
- I was advised by my classmates to do so
- I read the leaflet on the Department website and it attracted my attention
- Other

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/65zZxk1JCmDMqAaE8.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1SNuReIjXHjXqJYjkZRX5c3U2J8PB0LqJl6pTWheMNT8/edit?usp=sharing.

Appendix 3

Post-workshop and pre-experiment questionnaire for the first experiment

Post-workshop satisfaction questionnaire

The satisfaction questionnaire that follows is anonymous and is designed to receive general feedback on the respeaking workshop, and to investigate the participants' perception on the skills acquired, the difficulties encountered, and the level required at the beginning and end of the training. Some questions are posed on the expectations towards the upcoming experiment.

The respeaking workshop

Will you acquire the 3 ECTS for the workshop? Yes No

Overall, how much are you satisfied by the respeaking training experience?

1	2	3	4	5
(not at all)				(very much)

If you had the opportunity, would you continue with other workshops, courses, tutorials, etc., your respeaking training?

- I would like to.

- I do not think so.

Do you think that the introductory theoretical part of the training was adequate (well structured, complete, balanced) in order to then face practice at a later stage?

1 2 3 4 5

How much do you think the COVID-19 pandemic emergency and the resulting remote lessons have affected the quality and the enjoyment of your teaching experience?

1 2 3 4 5

What do you feel were the biggest obstacles for you in attending the workshop at a distance?

- Weak or intermittent Wi-Fi connection
- Disturbing or annoying elements at home (other people around, telephone, intercom, pets, etc.)
- Poorly performing PC compared to those I could have used at University
- Inappropriate technical equipment (headphones, microphone, software, or other)
- Little or less motivation to tackle the exercises by being physically not present in the class
- Other

To what extent do you feel that software, tools, equipment, instructions, technical support, etc. provided during the training were sufficient and/or helpful?

1 2 3 4 5

In addition to the theoretical part, the respeaking course took place over three months and was quite intensive. What do you think about the frequency for the lessons for each group?

- I would have preferred to take the course for an extra month, but with less hours per week.
- They were necessary and helpful in maintaining the right pace and monitoring progresses and my improvement curve.

Also based on the previous answers, and keeping in mind that future editions of the workshop will likely to be held in-person instead, would you recommend your peers to attend it?

1 2 3 4 5

If yes, could you please briefly explain why would you recommend others to follow this training?

Do you feel your knowledge of respeaking has improved after the training and do you feel you have gained new skills?

1 2 3 4 5

Could you indicate what skills you feel you have strengthened/improved or gained from scratch through the respeaking workshop?

In general terms, do you feel that the materials and exercises offered were in line with your level of language and preparation? Yes No More or less

In terms of length, density of terminology, and speech rate, to what extent do you feel that the materials and exercises offered met a progressive level of difficulty from the beginning to the end of the course?

1 2 3 4 5

<u>The experiment</u>

Have you taken part in a group for a research experiment before? Yes No

Having taken part in an experiment tied to a research project, do you think it was different/interesting/added value/overcommitted? If you can, briefly share your impressions on this.

Was the experiment (creation of teams, times, tools, methods, duration, etc.) presented and explained clearly enough?

1 2 3 4 5

Thinking about the upcoming experiment you will be a part of, do you feel adequately prepared to deal with it? Yes More yes than no More no than yes Yes

Thinking about the upcoming experiment you will be a part of, what are your feelings?

Anxiety Calm Concentration Fear Excitement Other

What expectations do you have on the level of difficulty of the videos that will be proposed for the experiment?

- It will be in line with practice during the workshop.
- It will be more difficult.
- It will be easier.
- I do not know.

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/5epNB9ansUWZjS757.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1eYLA6kTQZJkjtLN-Vr7aAc0CT2MZo-Nt4cWBsE7V9pA/edit?usp=sharing.

Accessibility

Prior to the respeaking course, I was informed about disability and Media Accessibility.

1 2 3 4 5

Before the course I knew what respeaking was. Yes No More or less

The respeaking workshop helped me to be more knowledgeable about hearing disabilities and media accessibility.

1 2 3 4 5

After the respeaking course my sensitivity and knowledge about media accessibility for the deaf and hard of hearing has increased.

1 2 3 4 5

In what ways do you feel you are more prepared and/or aware in terms of accessibility for the deaf and hard of hearing?
(More than one option can be selected).

- I don't feel more knowledgeable, nor more prepared on the subject.

- I am more aware and sensitive to the need for subtitling (live and otherwise) for audiovisual content.
- I continue to be interested in it through training or out of personal interest.
- I was struck by the dual accessibility for the deaf on one side and multilingual on the other.
- Other

If you want, you can further detail the answer(s) given above.

The additional specific questions in Italian n Media Accessibility and respeaking are available via Google Form at: https://forms.gle/7bbMznHxPLBuDXQd8.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1L2cqcAP7uPH_XtMLEp-gAgG_MYKUWhFQnTVzw7ZIkgo/edit?usp=sharing.

Appendix 4

Post-experiment questionnaire for the first experiment

Post-experiment questionnaire

The following questionnaire is useful to understand how the participants perceived the proposed videos in terms of level of difficulty and to detect if they encountered technical problems. We would also like to observe your impressions and/or feelings about the performance achieved. The questionnaire is not anonymous in order to link any obstacles in the workflow to those who report particular difficulties.

Name and surname:

Overall, how complex did you find respeaking experiment as a whole?

1 2 3 4 5

During the test, did you experience any technical problems? If yes, please specify them (slow software, locked PC, weak internet connection, audio interference, Dragon, etc.).

For intra and interlingual respeakers, did Dragon speech recognition software recognize the glossary of terms you added before the video began?

- Yes, all terms have been recognized
- No, I had to repeat several times most of the terms/I typed them with the keyboard
- Other

How was the length of the proposed video?

- Too long
- Too short
- Adequate

Did you know the topic of the video – environmental crisis, climate change, global warming?

- Very well
- Quite well
- Not really
- Not at all

Regarding speech rate, how complex did you find the test?

1 2 3 4 5

If you selected your answer in the range from 3 to 5, could you specify your difficulties (e.g., I missed some parts, I couldn't keep up with the speech, I summarized a lot, I introduced new information, etc.)?

Terminologically, how complex did the test seem to you?

1 2 3 4 5

As for the figures (quantities, percentages, dates, years, etc.), how do you feel you addressed them?

- I have understood and translated all or almost all of them
- I didn't understand many of them, but I could render some of them
- I generalized, giving quantitative or temporal references close or similar to the ST
- I understood none of them and rendered completely different references
- Other

In general, could you identify and summarize the main obstacles you encountered in this test (e.g., presence of many figures/dates/percentages, presence of proper nouns, complex and/or awkward topic, lack of clarity in the exposition, speaker's pronunciation, etc.)?

Overall, at the conclusion of the video, how did you feel performatively but also emotionally?

- Fully satisfied and confident
- I could have done much better, so a little dissatisfied
- I feel I did a lot of errors, so a little confused/unsatisfied.
- I haven't really been able to follow the ST, so I can't say.
- Other

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/GuWPmGJnyxtBM2id7.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1Djym45-pcNE3hJbyK1ei2iQLV3HUQDw6x1S0XKup0qM/edit?usp=sharing.

Appendix 5

Second experiment materials
(NTR analyses, video and audio recordings)

All analyses carried out for the five teams in the second experiment (EX2, Spanish to Italian) are available at the following cloud link: https://drive.google.com/drive/folders/1WuftlncfZ1pNA0MOiyk4Cut_sXbesroj?usp=sharing.

Folders are divided into Method 1, Method 2, Method 3, Method 4, and Method 5. For each method, three folders are available for the three source videos (Video 1, Video 2, Video 3) and they collect the audiovisual material recorded during the experiment: .mp3 recording, .mp4 videos (both source text and participant feeds, and with separated audio tracks). For each method also the NTR analysis template, with the transcription file is available in the dedicated folder 'Analysis'.

(Note that for Method 4, the folders of both participants who performed in this method are included).

Appendix 6

Pre-workshop questionnaire for the second experiment

The pre-workshop questionnaire for the second experiment was the same that was submitted to the first group of participants, with no or very little editions. For the detail on its structure and questions, refer to Appendix 2.

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/henh2ZMD9MhgUTVN8.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1PjMx4Vtvp1UgVElHO8cTGkuqF8uc8tlXB7gAw8nEBps/edit?usp=sharing.

Appendix 7

Post workshop and pre-experiment questionnaire for the second experiment

The post-workshop and pre-experiment questionnaire for the second experiment was the same that was submitted to the first group of participants, with no or very little editions. For the detail on its structure and questions, refer to Appendix 3.

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/naPhuWkP6rKquMD27.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/1PhFpWlBP7XbLPp0I4E9ETbe9MTV0TwxBosvk59J9JDw/edit?usp=sharing.

Appendix 8

Post-experiment questionnaire for the second experiment

The post-experiment questionnaire for the second experiment was the same that was submitted to the first group of participants, with no or very little editions. For the detail on its structure and questions, refer to Appendix 4 as the same questions were asked, only repeated for the three different source texts.

The Google Form original questionnaire in Italian, as it was sent to participants, is available at: https://forms.gle/3hH9yPnJvNTk92fq8.

Relevant responses are gathered and summarized in an Excel sheet at: https://docs.google.com/spreadsheets/d/16JQ7OvAqjhwhTvoLkcAjyJ_NCU8HFD9YSN86mhDcK58/edit?usp=sharing.

Appendix 9

Audio and video recordings authorization form

Oggetto:

LIBERATORIA PER LE REGISTRAZIONI AUDIO E VIDEO
LABORATORIO DI RESPEAKING

Io sottoscritto/a _____ nato/a _____ il _____,

AUTORIZZO

la dott.ssa Pagano Alice, dottoranda presso il Dipartimento di Lingue e Culture Moderne dell'Università degli Studi di Genova, a effettuare registrazioni audio e video, realizzate anche da Soggetti terzi autorizzati dalla medesima, nell'ambito del corso dedicato "Laboratorio di respeaking" per gli anni accademici 2020/21 e 2021/22. Acconsento altresì affinché le stesse vengano pubblicate, nella loro integrità o in modo parziale, nell'ambito di una tesi di ricerca di Dottorato in modo anonimo. Acconsento, infine, affinché i dati e le risposte raccolte tramite i questionari del laboratorio, anonimi e non, vengano pubblicati nella loro integrità o in modo parziale.
In fede,

Data e luogo