



This is a contribution from *Cultus: the Intercultural Journal of Mediation and Communication* 2019: 12

© **Iconesoft Edizioni Gruppo Radivo Holding**

This electronic file may not be altered in any way.

The author(s) of this article is /are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Training Respeakers: A Hybrid Profession for the (near) Future?

Annalisa Sandrelli

Università degli studi Internazionali - UNINT, Rome

Abstract

The digital revolution of the past twenty years has massively changed the language industry, now heavily technology-based. The use of technology has changed the pace and organisation of labour, but has also created new opportunities, by making it possible to provide new services. One such example is live subtitling via respeaking, a technique based on speech recognition technology. The skills and competences required to provide such services are still being investigated, with a view to informing training methods and curricula. The paper gives an overview of research activities in this field carried out at the Università degli Studi Internazionali - UNINT, Rome.

1. Introduction

The focus of this paper is on respeaking, a technique that straddles across the translation-interpreting divide. As the terms “translator” and “interpreter” are often used interchangeably in every-day language, it is perhaps useful to begin with a definition of both activities, provided by the International Association of Conference Interpreters (AIIC). The AIIC website points out that interpreting is a real-time service provided to participants (speakers and listeners) in a communicative event, while translators usually work in relative isolation with no immediate contact with target language readers. It then adds:

An interpreter works with spoken words in a particular context, conveying a message from one language to another, while translation

refers to the activity of transferring a written text from one language to another. [...]” (AIIC 2012)

This definition, albeit clear, is a little restrictive, as it does not explicitly include hybrid modes such as sight translation (spoken translation of written texts) and subtitling (written translation of spoken texts). This paper takes a wider perspective and talks about professionals working in the language industry, defined as follows:

‘Language industry’ covers professionals working in translation, interpreting, subtitling and dubbing, localisation, language technology tools development, international conference organisation, language teaching, linguistic consultancy. (LIND 2019)¹

One activity that has developed in the language industry in recent years is respeaking for live subtitling, a technique that relies on speech recognition technology and efficient human-machine interaction. Respeakers listen to live input and simultaneously reformulate it to speech recognition software that turns it into written subtitles. This method is widely used to provide live intralingual subtitles for the deaf and the hearing-impaired on television and in live events; the interlingual variant, which resembles simultaneous interpreting but produces a written target language text instead of a spoken one, has the potential to contribute to universal accessibility to information, culture and entertainment by bridging both linguistic and sensory barriers.

As respeaking shares elements of both translation and interpreting and is heavily technology-based, it makes sense to briefly summarise the main changes brought about by technological progress in the language industry and their impact on the training of translators and interpreters (§2). Then, the focus shifts onto respeaking itself, with a brief overview of relevant literature on research and training (§3) to serve as the background for a discussion of the respeaking-related teaching and research experiences at the Università degli Studi Internazionali di Roma, (§4). Finally, conclusions are presented in §5.

¹ The Language Industry Expert Group (LIND) provides consultancy services to the Directorate-General for Translation of the European Commission.

2. The technological turn in the language industry and in translator and interpreter training

In the first two decades of the 21st century the advent of the digital revolution has changed the language industry beyond all recognition. Of course, some language professionals rely on technology more than others, depending on their geographical location, specialisations, age, and so on; but at least in Western societies they all routinely use a variety of tools. In translation, the uptake of technology has been pervasive and very fast-paced. At the turn of the 20th century, translators were already expected to be able to use many different types of electronic tools (Austermühl 2001), but over the last couple of decades their range and efficiency have increased massively. Today, translators use online dictionaries, glossaries, term banks, corpora, CAT (Computer Assisted Translation) tools and, in many cases, machine translation (MT) software as well. This technological turn in translation has also created new tasks. In order to use MT systems efficiently, pre-editing and post-editing are necessary, and since the use of CAT tools has made it easier to split large jobs among several translators, tasks such as project management, TM (translation memory) management and revision have acquired growing importance. Thus, technology has had a huge impact on the pace and organisation of labour. The latest trend is the shift to cloud-based tools, which make it easier to optimise workflows and perform quality checks on large translation projects. Moreover, as turnaround times are significantly lower than in the past, some translators have been experimenting with speech recognition software to dictate their translations and increase productivity. In short, today translators need to be able to use a variety of tools to lever competitive advantage on the market (Rothwell and Svoboda 2019).

The technological turn has affected the interpreting profession more recently and to a lesser extent, but it is now definitely here to stay. Fantinuoli (2018b) classifies the impact of technology on interpreting into three areas: Computer Assisted Interpreting (CAI), Remote Interpreting (RI) and Machine Interpreting (MI), the latter still in its early stages of development.² Interpreters use a wide range of online and off-line tools (term bases, glossaries, “do-it-yourself” corpora, and so on) before and after assignments for preparation

² Readers are referred to Fantinuoli (2018a) for a fuller overview.

purposes and for continuous professional development. Moreover, there are dedicated terminology management tools such as *Interplex* to support interpreters with their glossaries. An even more sophisticated tool, *Interpretbank*, has been expressly designed for use in the booth (optimised for quick term look-up during simultaneous interpreting assignments), and features the integration of Automatic Speech Recognition (ASR) technology to transcribe terms and figures and prompt interpreters with suggestions.³ In addition, remote interpreting (with the interpreter away from the primary participants' location) and videoconference interpreting (with participants distributed across several sites) are becoming more and more common; and there are several interpreting apps that make it possible for clients to listen to the interpretation via their own devices (smartphones, tablets, laptops, and so on) instead of hiring specialised equipment.⁴

All the above developments have had an impact of the training of translators and interpreters, where technology is acquiring growing importance. For example, the European Master's in Translation (EMT) network⁵ includes technology competences in its core curriculum: search engines, corpus-based tools, CAT and MT tools, corpora, as well as the ability to handle different file formats and the localisation of multimedia texts such as websites and videogames. The growing importance of technology training in EMT institutions has been confirmed in a recent survey (Rothwell and Svoboda 2019). It found that almost 100% of programmes in the network now teach the use of translation memories and terminology tools and that a majority also teach MT and post-editing (71%) and cloud-based tools (69%). Finally, 35% also offer training in subtitling, which in some cases includes respeaking. In addition, EMT members seem to be moving away from a "traditional" view of translation as an activity involving exclusively written texts, for example by embracing multimedia translation.

Of course, the EMT network is a select group of institutions and cannot be considered representative of all the academic programmes

³ See *Interplex* (<http://www.fourwillows.com/interplex.html>) and *Interpretbank* (<http://interpretbank.com/>).

⁴ See for example *Interprefy* (<https://interprefy.com/>), *Interactio* (<https://interactio.io/>), *Linguali* (<https://linguali.com/en/>) and many more.

⁵ The European Master's in Translation (EMT) is a network of universities offering MA degree courses in translation, developed under the aegis of the Directorate-General for Translation of the European Commission. See https://ec.europa.eu/info/resources-partners/european-masters-translation-emt_en

in this field. However, the growing interest in technology on the part of translator and interpreter training institutions has also been confirmed by the 2019 edition of the LIND survey of the European Language Industry, whose respondents include professionals, language services companies and training institutes. One aspect that is especially relevant for the purposes of this paper is that speech recognition technology has gradually been adopted in both professional translation and interpreting, so much so that it is contributing to blur the boundaries between the two. Several CAT tools already include speech recognition features to enable translators to work using their voice;⁶ in addition, speech recognition technology can be used in simultaneous interpreting (see the above-mentioned *InterpretBank* tool) and in subtitling (live and recorded). The popularity of speech recognition has been noted by translator and interpreter training institutions:

Training institutes show a clear interest in dictation and subtitling technology. It is also interesting to note that dictation technology scores almost as high as machine translation with independent professionals. (LIND 2019)

It is now time to focus on speech recognition and on its application to respeaking.

3. Respeaking: an overview

Today, respeaking is the method of choice in many countries to provide accessibility services on television. It is also used in live events, such as conferences, lectures, and so on.⁷ As was mentioned earlier, it relies on speech recognition technology, originally developed in the 1990s for dictation purposes. A clarification is in order here: there are two types of speech recognition systems, namely speaker-dependent and speaker-independent ones. The former can transcribe any speaker's words with reasonable accuracy. No prior preparation of the software is required, provided the input is of good quality (i.e. no overlapping voices, foreign accents or other sources of signal noise).

⁶ See for example MateCat (<https://www.matecat.com/>) and Hey memoq (<https://www.memoq.com/products/hey-memoq>).

⁷ In Italy the preferred method for live subtitling is still stenotyping. Respeaking is used on television and in live events, but to a much lesser degree.

In contrast, speaker-dependent recognition programmes require users to create a voice profile and to train the system by reading aloud some texts. As a result, the software gradually becomes familiar with the user's voice, pronunciation and speech patterns. This latter type of speech recognition technology is the one used for live subtitling via respeaking:

[...] 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 (DHOH) audience, to a speech recognition software, which turns the recognised utterances into subtitles displayed on the screen with the shortest possible delay. (Romero-Fresco 2011: 1)

The live subtitles produced via respeaking are used not only by deaf people, but also by non-native speakers and students (who may wish to listen and read at the same time for language learning purposes), and by the general public whenever it is impractical to listen to sound (e.g. at the gym or in an underground station). There is also an interlingual variant of respeaking (i.e. between two languages), which is a cross-over between simultaneous interpreting and interlingual subtitling, depending on whether you look at it from the point of view of the process or the product:

With regard to the process, 'interlingual respeaking' [...] is really a form of simultaneous interpreting, while the product, in the intralingual as well as the interlingual variant, is a set of subtitles. (Romero-Fresco and Pöchhacker 2017: 158)

Interlingual respeaking is still in its infancy, but it is potentially a very inclusive translation mode, as it can be used to make multilingual content accessible to foreign and hearing-impaired subtitle users at the same time. It is worth mentioning that in April 2019 the European Commission adopted the European Accessibility Act, a directive aimed at facilitating disabled people's access to information and entertainment: this measure is expected to boost services such as subtitles for the deaf and hard of hearing, audio-description, spoken subtitles and sign language interpretation. Moreover, while in the past the audiovisual sector included only traditional media such as cinema and television, today a key role is also played by Internet-based streaming platforms, the gaming industry and the emerging Virtual Reality productions. All these new materials need to be made

accessible not only across sensory barriers (intralingual) but also across languages and cultures (interlingual). Finally, there is an untapped market in the live subtitling of on-line radio broadcasts, in the live subtitling of online classes (MOOC) and webcasts. Thus, it would seem that demand for this kind of service (in both the intralingual and interlingual variants) is likely to increase in the near future, and that it can provide plentiful employment opportunities to graduates with the necessary knowledge, skills and competences.

Clearly, in order to design respeaker training courses effectively, it is important to understand the process and skills involved. Over the last 10-15 years, research into respeaking has been developing steadily; initially only on the intralingual variant, and in the last 5 years on the intralingual one as well. The studies produced until now have been either descriptive-empirical (i.e. case studies analysing respeaking services in specific settings) or experimental (i.e. tasks carried out in a laboratory). A brief summary of key studies is provided in §3.1, while a quick overview of methods and approaches used in the training of respeakers is given in §3.2.

3.1 Respeaking research

As respeaking is such a young translation mode, research is still scanty. The available literature focuses on three main topics: the skills and competences required for intra- and inter-lingual respeaking; the accuracy of the subtitles produced via respeaking; and the viability of respeaking as a service, i.e. how to deliver it in various settings whilst ensuring the best possible quality.

As regards the process, a number of studies have highlighted the similarities between simultaneous interpreting and respeaking (Marsh 2004; Eugeni 2008). Recent studies have tried to determine whether a background in interpreting or subtitling can confer an advantage in the acquisition of respeaking skills and competences (Szarkowska *et al.* 2018, Dawson 2018). Early results seem to indicate that an interpreting background can help, but is not mandatory, and that subtitling skills are also needed to perform respeaking satisfactorily.

Assessing the accuracy of live subtitles is another key topic being investigated. Starting with intralingual respeaking, several models have been proposed, but the most widely used one is the NER model (Romero-Fresco 2011, Romero-Fresco and Martínez 2015). It distinguishes between (software-related) recognition errors and (human) edition errors, and attributes a score to each error depending

on its severity (minor, standard or serious), calculated in terms of the effect on viewers. The NER model has become established as the industry standard and an accuracy score of 98% has been suggested as the minimum quality threshold for (intralingual) live subtitles (Romero-Fresco 2011). As regards interlingual respaking, Romero-Fresco and Pöchhacker (2017) have recently developed the NTR model, which once again distinguishes between recognition errors and human (translation) errors: the latter include content-related errors (omissions, additions and substitutions) and form-related errors (affecting grammatical correctness and style). The model needs to be validated in various settings to determine whether a 98% accuracy rate is a feasible benchmark in interlingual respaking too.

Of course, other aspects also play a role in the service delivered to users. One of them is the average latency of subtitles, caused by the combination of the respeaker's time lag in relation to the original speaker and the delay caused by data processing by the speech recognition software. This can become a problem in conferences when slides are used or on television in programmes with visual information (graphics, maps, and so on): if there is too much delay, the subtitle is displayed when the relevant visuals have already disappeared (Romero-Fresco 2011). In addition, the speaking speed of the original speaker is another key factor, because respeakers have to add punctuation marks orally. This means that, in order to produce verbatim subtitles, they have to say more words than the original speaker. In most cases, therefore, respeakers are forced to edit the subtitles in order to keep up with the original speaker, by finding succinct ways of expressing the same ideas. This is where both interpreting and subtitling experiences may come in handy (Romero-Fresco 2009). Another important aspect is how subtitles are displayed, whether as scrolling continuous text (letter by letter, syllable by syllable or word by word) or in blocks of text, and in one, two or three lines at a time. Moreover, on television subtitles may be provided as either open or closed captions, while in live events they may be projected onto a screen or relayed to the users' personal devices (smartphones, tablets or laptops) via a network connection. Furthermore, the live subtitling service may be provided by two respeakers taking it in turns to work (just like two interpreters in the booth), or by a team which includes other professionals as well, such as a subtitle editor, a proof-reader and others. All the possible variations in the set-up of a live subtitling service shape the service itself and make direct comparisons between different settings difficult. That is why there is a need for more

empirical data from actual respeaking settings, even if they are relatively small-scale events. For example, Marchionne (2010-11) reported on an experiment in live subtitling via interlingual respeaking on television (in the French-Italian language combination), while Serafini (2014-15) reported on an interlingual respeaking service (English- Italian) at a small film festival in Italy. A more comprehensive research effort of this kind is the work being carried out by Moores (2018) on the feasibility of respeaking in museum tours, conferences, lectures and QandA panels after cinema screenings and theatre shows. The aim, in all these cases, is produce best practice guidelines to organise such services efficiently.

As this brief overview shows, the actual skills and competences required to provide such “hybrid” language services are still being investigated; nevertheless, a few universities have begun to train respeakers on the basis of what has emerged until now.

3.2 Training respeakers

Just over ten years ago, the first intralingual respeaking modules began to be offered by the University of Antwerp (Belgium) in its MA in Interpreting; and by the Universitat Autònoma de Barcelona (Spain), and by Roehampton University (UK) in their MAs in Audiovisual Translation (Romero-Fresco 2011). All of these courses are focused primarily on intralingual respeaking and are aimed at providing students with the skills they need to work in live subtitling for the deaf and hearing-impaired, mostly on television.

Respeaking modules generally include an introduction to the speech recognition software used in the course, which for widely spoken languages such as English, Spanish and German tends to be Nuance’s *Dragon Naturally Speaking*. Romero-Fresco (2011, 2012) proposed a taxonomy of respeaking skills to form the backbone of training courses. The taxonomy distinguishes between the skills needed in the preparation phase (before the process), during respeaking proper and after the task. Moreover, the classification highlights which skills are shared with subtitling, which ones with simultaneous interpreting and which ones are specific to respeaking:

Learning to respeak involves the acquisition of a number of unique skills, including (i) how to listen, (re)speak, watch (the screen) and type [...]; (ii) how to keep an optimum rhythm and produce valid respeaking units that display appropriately within the space allocated

for the subtitles on the screen; and (iii) how to edit the speech in real time, when it is not possible to keep up with the speech delivery rate. (Romero-Fresco 2012: 99).

Initially, students need to create their voice profile and become familiar with the features of the speech recognition software. They gradually learn how to speak to ensure maximum accuracy; to ensure seamless transcription, students learn to speak relatively normally, though maintaining a steady pace, avoiding disfluencies (hesitations, mispronounced words, and so on) and articulating clearly. As they become more confident in dictating to the software, they learn how to improve recognition by adding new words to the software's Vocabulary Editor. To do so, students will research the topic, prepare a glossary and then import it into the programme. They will also learn to anticipate recognition errors by, for example, avoiding the use of foreign words or neologisms.

The main similarities between respeaking and simultaneous interpreting come to the fore during the process. Just like interpreting students, respeaking trainees must develop concentrated listening, text analysis and reformulation skills, as well as the ability to listen and speak at the same time. Condensation skills are also important, in order to keep up with the pace imposed by the original speaker. Stress management and monitoring skills are especially important too. However, multitasking is even more complex in respeaking than in simultaneous interpreting, because there is an additional effort involved: not only listening and speaking, but also reading and writing (to correct any errors in the subtitles). Another skill that characterises respeaking is the ability to mentally visualise the original speaker's sentences in a written form and decide how to edit them, how to segment the text and where to insert punctuation marks orally, i.e. by dictating punctuation to the speech recognition software. In addition, it may be necessary to change the colour of the subtitles or to use different labels to identify different speakers for the benefit of deaf subtitle users. Finally, after the process, it is important to be able to assess one's performance and the quality of the subtitles, and identify strengths and weaknesses to ensure constant progress. In this sense, the NER model is also used as a diagnostic tool in training courses. If there is a pattern of misrecognising words, the trainee will need to improve his/her voice model and dictation skills; while if there are more content-related errors, the trainee must work on his/her text analysis and reformulation skills to produce a more accurate rendition.

As this brief description shows, respeaking is an extremely complex activity in which all the language competencies (listening, speaking, reading and writing) are involved, owing to the diamesic shift from a spoken input to a written end-product. Moreover, efficient human-machine interaction is required, because the respeaker actually produces an intermediate spoken text aimed at the speech recognition software (not the audience), which then converts it to a written text for the users. When respeaking is performed interlingually, the translation difficulty is added to the equation. To date, only one training course for interlingual respeaking has been organised: this was a four-week pilot training course delivered online within the ILSA (*Interlingual Live Subtitling for Access*) project (Dawson 2018). 50 students with a training background in subtitling or interpreting and with an English-Spanish language combination participated in the course. At the end of the course, over 40% of students managed to achieve an NTR score of (or over) 98%, thus proving that interlingual respeaking is indeed feasible. Moreover, the student interpreters performed better than the subtitlers on average, though some subtitlers did well too.

As this overview of respeaking research and training has shown, there is still a need for both empirical and experimental data regarding this practice, if it is to become more established. Section 4 reports on the contributions to the field coming from the UNINT in Rome.

4. Respeaking at UNINT: teaching and research

The Faculty of Interpreting and Translation of UNINT introduced an intralingual Respeaking Workshop in 2014. It is a 30-hour option offered to interpreting students in the second year of their MA in Interpreting and Translation. Owing to its short duration, the module is only meant as an introduction to speech recognition technology and intralingual respeaking for television and live events.

The module runs in Semester 2, when students have already been exposed to at least one semester of consecutive and simultaneous interpreting training: this ensures that they are able to listen and speak at the same time, summarise an idea or paraphrase a concept, and they have relatively fast reaction times. Some of them have also had an introduction to audiovisual translation and subtitling and are therefore relatively familiar with subtitling conventions, as well as with working on written text (including punctuation). Despite this, initially most trainees struggle with dictation, text segmentation and adding oral

punctuation: many of them find it difficult to articulate clearly, to keep a steady pace, to pause at the right time and to avoid using the prosodic resources they would normally use when interpreting. Over time, they learn how to work with the software to improve recognition, by adopting a flat, slightly robotic intonation, and by paraphrasing words that are not likely to be in the vocabulary of the programme.

Classes take place every week in a computer lab where the speech recognition software (*Dragon Naturally Speaking*) is available on all the machines. Students also have a weekly 3-hour timetabled autonomous practice session and can also use the lab on an open-access basis whenever the room is free. The course is hands-on and intensive, so class size has been kept deliberately small, with a maximum of 15 students per group. Over the course of the semester, students are given an overview of the needs of the deaf and of the hearing-impaired, to make sure they understand how to produce readable subtitles. They are also expected to carry out some independent research into the topics assigned for practice and to prepare the software for respeaking tasks by adding new words to the vocabulary: this teaches them how to prepare for respeaking assignments.

During the course, students are exposed to a variety of spoken texts presenting different degrees of difficulty: they start with narrations and slow conference speeches, to move on to sports broadcasts, weather forecasts, news programmes, sports, interviews and talk shows. Moreover, students progress from mono-thematic genres and constrained vocabulary with the involvement of one slow speaker, to talk-shows and debates involving many speakers and open-ended topics. Indeed, the final exam consists of two practical respeaking tasks involving a monologic and a dialogic text (such as an interview).

Evaluation and self-assessment methods are also discussed and exemplified at length in class, to ensure that students develop the tools to assess their own output and identify strengths and weaknesses. In this sense, they are encouraged to apply the NER model to their own work to measure subtitle accuracy

It must be highlighted that, although the module cannot be said to produce professional respeakers, over the last few years a significant number of students (at least 3 or 4 in every group) have been offered placements or jobs in this field, while several others have reported using respeaking skills for other purposes, including to subtitle pre-recorded programmes, to carry out translation work and to transcribe meeting recordings.

In addition, the development of this training module has taken place in parallel with respeaking research at UNINT, in two ways: a number of small-scale case studies carried out as part of students' MA dissertations and a larger experimental study for projects SMART (*Shaping Multilingual Access with Respeaking Technology*) and SMART 2 on interlingual respeaking.

The case studies reported below can roughly be classified as empirical and observational studies (aimed at collecting and analysing authentic respeaking data in various settings) or as experimental studies. In all cases, the methodology and tools for data collection were developed jointly by the students and their supervisor (the present author), to ensure not only the validity of the studies, but also data comparability and consistency. The first two studies (in chronological order) focused on an analysis of respeaking practices in two television settings: the first one was a descriptive study on respeaking at SWISS TXT, the company providing subtitles to SRF, the German-language channel of the Swiss state broadcaster (Monsorno 2010-2011); the second one focused on respeaking at RAI, the Italian public broadcaster (Astuto 2014-2015).

Monsorno's study investigated the respeakers' educational and professional backgrounds by means of a dedicated survey questionnaire, and found that they were all young graduates from ZHAW (the Translation and Interpreting department at Winterthur, near Zurich) with some professional translation and interpreting experience. The accuracy of the subtitles produced by the respeakers was analysed using the NER model. The corpus included the weather forecast programmes broadcast in a given week, and the specific focus of the analysis was on the text condensation strategies used to cope with speed and information density. The MA student's case study was then implemental in defining the respeaking challenges posed by a specific TV genre, as was illustrated in Sandrelli (2013).

A few years later, Astuto (2014-2015) shifted the focus to Italy. As respeaking is not a well-established practice here, a survey of the respeaking market was carried out by means of a questionnaire sent out to the main stakeholders and companies working in the field, to shed light on the training and professional experiences of respeakers. The results of the survey showed that five years ago speech recognition technology was still used mostly for reporting purposes (for example in court hearings and company meetings), while it was not common in live subtitling. As RAI was the only broadcaster to make (limited) use

of respeaking, contacts were made to visit the subtitling unit there and observe the subtitlers at work.

Thanks to RAI's willingness to collaborate, it was possible to obtain the video-recordings of all the TV news programmes broadcast on *RAI Uno* for a week. At the time, the set-up was as follows: some TV news programmes were subtitled by stenotypists (working in-house), while the other editions were subtitled via respeakers working remotely. A NER analysis was carried out on the data, which revealed that while the highest accuracy rate was achieved via stenotyping (99.5%), the respoken subtitles reached the 98% quality threshold (see §3.1) in 3 out of the 4 editions of the TV news in the corpus. Both methods, therefore, seemed to ensure high quality subtitles; however, they differed greatly in terms of latency, which in stenotyping ranged between 1 and 7 seconds, as opposed to the 6-35 seconds in respeaking. The huge variability in latency in the respoken subtitles was caused by signal delay and technical problems related to the remote working set-up. Consequently, these values cannot be considered representative of actual differences in latency between stenotyping and respeaking, but merely as indications that during that specific week there were technical criticalities. This partially explains why the use of this technique is still lagging behind in Italy, as it is considered less reliable than the tried and tested stenotyping method.

Moving on to interlingual respeaking, another empirical study (Luppino 2016-2017) compared interlingual respeaking and simultaneous interpreting in the same conference setting. A fully accessible conference took place at UNINT in June 2015, where both simultaneous interpreting and interlingual respeaking (English-Italian) were offered to the audience. The first objective of the study was to create a multimedia corpus of the entire conference, featuring video clips of the speakers, audio clips of the interpreters and the respeakers, transcripts of the original speeches and of the interpreted output, and of course the interlingual subtitles produced via respeaking. Then, a smaller sub-corpus of 4 speeches (2 read and 2 delivered impromptu) was selected to assess how much of the semantic content of the source language speeches was conveyed to the audience via the interpreted speeches and via the subtitles.

First of all, the NRT model was applied to evaluate accuracy in a conference setting. Unfortunately, Luppino's findings show that the scores were far off the 98% mark on all 4 speeches. The best scores were obtained on the two read presentations; this may be explained by the fact that the respeakers received the presentations in advance and

were able to prepare the software by adding the relevant vocabulary. This benefit is especially evident with the low number of recognition and form-related errors, which indicates good human-machine interaction.

A dedicated analysis grid was also developed and applied to the data. Results show that the main difference between the simultaneously interpreted speeches and the target language subtitles was the quantity of omitted information, which was greater in interlingual respeaking. However, in terms of factual errors (distorted information) there was no significant difference between the two. In a sense, this could be taken as an indication that the two modes could co-exist side by side, leaving it up to the participants to decide which service(s) to use. The results of the study also point to the need to work on trainee respeakers' condensation skills, to ensure they are able to omit redundant items and reduce the target language text strategically (see Sandrelli forthcoming).

The same accessible conference was used as a source of material for another case-study, this time on intralingual respeaking. Capomaccio (2016-2017) transcribed the final conference roundtable that was held in Italian. A short product presentation was selected to organise an experiment involving 12 respeaking trainees at the end of their 30-hour respeaking course. The aim of the study was to compare the performance of the professional respeaker at work on the day with the performance of the trainees, not only in terms of accuracy rates (by applying the NER model), but also in terms of processing strategies. As it was a simulation of a respeaking assignment, participants were sent background material on the product and the speaker in advance for preparation purposes. On the day of the experiment the trainees were given a list of words to be added to the Vocabulary of the speech recognition programme; then, they were asked to respeak the speech that had been divided into two video clips. After subtitling each part, they filled in a questionnaire with comprehension questions and questions on their strategies for dealing with specific challenges (e.g. the presence of English words, ungrammatical sentences, and so on).

Unsurprisingly, the professional respeaker achieved a higher accuracy rate (around 96%), though half of the students achieved rates over 90% after 30 hours of training. The main sources of error in their output were recognition and omissions. 'Recognition errors' can be tackled by working more with the speech recognition software, given that performance improves over time. 'Omissions', on the other hand,

can be improved in training through specific tasks that develop the ability to omit strategically, through selecting redundant items and paraphrasing text to make it more succinct.

All of the above studies have inherent limitations in their size and specific settings, which make it impossible to generalise their findings. However, they do highlight some trends that are possibly worthy of more investigation. In this sense, this overview of respeaking research at UNINT ends with a brief discussion of 2 pilot projects on interlingual respeaking, involving 26 students from UNINT, Surrey and Roehampton Universities.⁸ The experiments focussed on identifying the skillset that would best facilitate the acquisition of interlingual respeaking skills. The findings of the projects have been reported in various conferences (for example in Sandrelli *et al.* 2019) and will be the object of a future publication (Davitti and Sandrelli *forthcoming*), so they are only briefly mentioned here for the sake of completeness.

The projects used a variety of data collection methods and tools: questionnaires, respeaking data, screencast recordings, and self-reflection TAP (think-aloud-protocol) sessions. The participants included students with training backgrounds in subtitling, dialogue interpreting, consecutive interpreting, simultaneous interpreting and intralingual respeaking, or a combination of two or more of these. To prepare them for the respeaking experiment, a “crash course” (6-8 hours) in interlingual respeaking was delivered face-to-face in the three sites. At the end of the crash course, two interlingual respeaking tests were collected from each student. The NTR model was applied to all the data, and data from questionnaires and TAP sessions were also analysed.

The results of the project can be summarised as follows: the best performers were those students with a composite set of skills comprising interpreting, subtitling and intralingual respeaking. Although interpreting trainees are already equipped with some of the required skills, many of them struggled with multitasking and monitoring written output. In contrast, the subtitling trainees are used to translating from spoken to written language, but found it hard to

⁸ SMART (*Shaping Multilingual Access with Respeaking Technology*) was funded by Surrey University between January and May 2018 and saw the participation of UNINT and the University of Vigo; this was followed by SMART 2 (funded by UNINT and with the same partner institutions) from September 2018 to September 2019.

cope with the real-time element of interlingual respeaking. Once again, the size of this pilot project makes it impossible to draw definitive conclusions, but it indicates some interesting patterns that certainly deserve to be investigated further.

5. Conclusions

This paper has given an overview of (intra)lingual and interlingual respeaking for live subtitling, a hybrid translation mode that requires very sophisticated language and translation competencies, the ability to use speech recognition software efficiently and to monitor one's output. As such, it involves complex multitasking. In this sense, it is certainly a very demanding task. Research has highlighted that variability in working set-ups has a direct influence on how respeaking is performed and the quality levels it is possible to achieve. What is still lacking is some standardisation and best practice guidelines in professional respeaking.

At the same time, as demand for both intra)lingual and interlingual respeakers is set to increase, there is a need to expand and improve existing training opportunities. Research seems to indicate that, while a background in interpreting may be an advantage, subtitlers can also become proficient respeakers. However, as respeaking requires a variety of skills, a modular approach to training may be the best option, enabling subtitlers to focus on improving their "live performance" skills, whilst allowing interpreters to concentrate on subtitle segmentation, oral punctuation and working with written text. In this sense, there is a demand for better training to be offered not only in academic programmes, but also via bespoke company training. Companies that are already active in the audiovisual sector are likely to need "top up" training courses for their existing staff. Before this can happen, however, there is a need for more research into the acquisition of respeaking skills. This is what the new, full-scale SMART project, based on the results of the pilot, is set to achieve in the years to come, starting in early 2020.⁹

It is hoped that this paper has shown that there is a huge potential in new translation-related professional tasks based on speech recognition technology. Professionals and academia must lead the

⁹ The new project will be funded by the UK's Economic and Social Research Council and will be led by Surrey University.

dialogue on these issues in collaboration with the technology industry, to ensure that progress is technology-based, not technology-driven.

References

- Astuto, A. 2014-2015. *Il respeaking in Italia: analisi di un genere televisivo*, Unpublished MA dissertation, UNINT, Rome.
- Austermühl, F. 2001. *Electronic Tools for Translators*. Manchester: St Jerome.
- Capomaccio, C. 2016-2017. *Il respeaking intralinguistico di conferenza: uno studio sperimentale sulle strategie di elaborazione del testo*. Unpublished MA dissertation, UNINT, Rome.
- Davitti, E. and Sandrelli, A. (forthcoming). “Embracing the complexity: findings from a mixed-method study on interlingual respeaking”. Special issue of *JAT- Journal of Audiovisual Translation*.
- Dawson, H. 2018. “Interlingual respeaking: a short online course”, Paper presented at the *6th International Symposium on Accessibility and Live Subtitling*. Milan, 14 September 2018.
- Eugeni, C. 2008. “A sociolinguistic approach to real-time subtitling: respeaking vs. shadowing and simultaneous interpreting”. In C. J. Kellett Bidoli and E. Ochse (eds) *English in International Deaf Communication, Linguistic Insights*, Bern: Peter Lang, 357-382.
- Fantinuoli, C. (ed.) 2018a. *Interpreting and Technology*, Berlin: Language Science Press.
- Fantinuoli, C. 2018b. “Interpreting and technology. The upcoming technological turn”. In C. Fantinuoli (ed.), *Interpreting and Technology*, Berlin: Language Science Press, 1–12.
- Luppino, D. 2016-17. *Interpretazione simultanea e respeaking interlinguistico: due modalità di traduzione a confronto. Un contributo alla ricerca*, Unpublished MA dissertation, UNINT, Rome.
- Marchionne, F. 2010-11. *Il respeaking interlinguistico. Sperimentazioni per una evoluzione della sottotitolazione in diretta*, Unpublished MA dissertation, University of Macerata, Italy.
- Marsh, A. 2004. *Simultaneous Interpreting and Respeaking: a Comparison*, Unpublished MA dissertation, University of Westminster, UK.
- Monsorno, S. 2010-2011. *Il respeaking in televisione: strategie di condensazione e valutazione del tasso di accuratezza dei sottotitoli*. Unpublished MA dissertation, UNINT, Rome.

-
- Moores, Z. 2018. "Respeaking at live events: ensuring quality in diverse settings". Paper presented at the *6th International Symposium on Accessibility and Live Subtitling*, Milan, 14 September 2018.
- Romero-Fresco, P. and J. Martínez 2015. "Accuracy rate in live subtitling- the NER model". In J. Díaz-Cintas and R. Baños Piñero (eds), *Audiovisual Translation in a Global Context. Mapping an Ever-changing Landscape*, London: Palgrave, pp. 28-50. Url: <http://www.captiontest.com/roehampton%20NER-English.pdf>
- Romero-Fresco, P. and F. Pöchhacker. 2017. "Quality assessment in interlingual live subtitling: The NTR Model". *Linguistica Antverpiensia New Series: Themes in Translation Studies* 16, 149-167.
- Romero-Fresco, P. 2009. "More haste less speed: edited vs. verbatim respeaking". *Vigo International Journal of Applied Linguistics (VIAL)* 6, 109-133. URL: <http://vialjournal.webs.uvigo.es/pdf/Vial-2009-Article6.pdf>
- Romero-Fresco, P. 2011. *Subtitling through Speech Recognition: Respeaking*. Manchester: St. Jerome.
- Romero-Fresco, P. 2012. "Respeaking in translator training curricula. Present and future prospects". *The Interpreter and Translator Trainer* 6(1), 91-112.
- Rothwell, A. and T. Svoboda. 2019. "Tracking translator training in tools and technologies: findings of the EMT survey 2017". *JosTrans- The Journal of Specialised Translation* 32, 26-60.
- Sandrelli, A. 2013. "Reduction strategies and accuracy rate in live subtitling of weather forecasts: a case study", Paper presented at the 4th International Symposium on Live Subtitling: *Live Subtitling with Respeaking and Other Respeaking Applications*. Barcelona, 12 March 2013.
- Sandrelli, A. *forthcoming*. "Interlingual respeaking and simultaneous interpreting in a conference setting: a comparison". *Intralinea*.
- Sandrelli, A., Davitti, E., and P. Romero-Fresco 2019. "Interlingual respeaking (IRSP): triangulating quantitative and qualitative data across different subject groups", Paper presented at the International Conference *Media for All 8. Complex Understandings*. Stockholm, 17-19 June 2019.
- Serafini, G. 2014-2015. *La sottotitolazione interlinguistica in tempo reale: il caso Ericsson Olanda e l'esperimento presso il festival Sediccicorto 2014*, Unpublished MA dissertation, University of Bologna.
- Szarkowska, A., Krejtz, K., Dutka, Ł. and O. Pilipczuk (2018) "Are interpreters better respeakers?". *The Interpreter and Translator Trainer* 12 (2), 207-226.

Web References

- AIIC 2012. “Interpreting is not translating”. *aiic.net*, April 23, 2012. Url: <http://aiic.net/p/4002>
- EMT 2017. European Master’s in Translation. *Competence Framework 2017*. Url: https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf
- LIND 2019. *European Language Industry Survey* https://ec.europa.eu/info/departments/translation/language-industry-platform-lind_en#documents
- EC 2019. *Directive 2019/882 of the European Parliament and of the Council on the accessibility requirements for products and services*, 17 April 2019. Url: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0882&qid=1564591292135&from=EN#d1e32-100-1>