A Practical Proposal for the Training of Respeakers[1]

Marta Arumí Ribas, Universitat Autònoma de Barcelona
Pablo Romero Fresco, Heriot-Watt University

ABSTRACT

In the field of Audiovisual Translation, some disciplines still have a long way to go in terms of visibility. Speech recognition-based subtitling, also known as respeaking, is a case in point. Even though it seems to be consolidating as the preferred method of providing intralingual live subtitles for the Deaf and Hard-of-Hearing in many TV channels, it is far from being consolidated regarding research and especially teaching.

Building on the research carried out so far in the field, the present article attempts to tackle the question of the training of respeakers. First of all, respeaking is presented, described and compared to subtitling and interpreting. Then, a full account is given of the skills required for a respeaker, whether they are to be obtained from subtitling, interpreting or specifically from respeaking. Finally, a practical proposal for the training of respeakers is put forward by way of practical exercises geared at providing students with the required skills.

KEYWORDS

respeaking, subtitling, simultaneous interpreting, training, skills and competences.

1. Introduction

Much has been said and written about the invisibility of translation in the past years, which, not surprisingly, has led to an increasing recognition of translators as professionals and Translation Studies as an academic field. Audiovisual Translation (AVT) seems to have followed suit, going in this case hand in hand with the ever-important area of Media Accessibility. Yet, some disciplines within AVT still have a long way to go in terms of visibility. Speech recognition-based subtitling, also known as respeaking, is a case in point.

This does not mean, however, that respeaking is still to be established as a professional activity. On the contrary – the BBC, for example, has been using this method to provide live subtitles for the Deaf and Hard-of-Hearing for the past seven years. Most importantly, the growing demand for this type of subtitles, as determined both by European and national legislation, has led many broadcasters to choose respeaking over keyboards or stenotyping as a more cost-effective alternative. Yet, this incipient consolidation of respeaking as a subtitling technique has not been reflected in the areas of teaching and research. Whereas the former is practically non-existent, which explains why broadcasters are currently having to train their own professionals from scratch, research is still at its very beginning,

concerned with general descriptions of the discipline (Eugeni 2006, Romero Fresco 2008) and its different national practices (Orero 2006, Remael and van der Veer 2006). Some attention has also been given to its position within the general framework of Translation Studies by comparison/opposition to similar disciplines such as subtitling and interpreting (Eugeni 2008). Yet, except for a few mentions in passing, only van der Veer (2007) has looked at respeaking from the point of view of teaching, which Baaring views as "a really interesting question, both from a research and a didactical point of view and one that should be investigated further" (2006).

Building on the research carried out so far in the field, the present article attempts to tackle this question of the training of respeakers. Firstly, respeaking is presented, described and compared to subtitling and interpreting, drawing on the relevant literature available. Then, a full account is given of the skills required for a respeaker, whether they are to be obtained from subtitling, interpreting or specifically from respeaking. Finally, a practical proposal for the training of respeakers is put forward by way of practical exercises geared at providing students with the required skills.

2. Defining respeaking

One of the consequences of the very little research carried out so far in this field is the lack of established terminology to refer not only to the professionals engaged in this discipline but to the discipline itself. A quick look at some of the publications available yields several long and precise labels such as *speech-based live subtitling* (Lambourne et al. 2004), (*real time) speech recognition-based subtitling* and *real-time subtitling via speech recognition* (Eugeni 2008). Shorter and perhaps more functional alternatives are *revoicing* (Muzii 2006), *voice-writing* (Vincent 2007) and *real-time voice-writing* (Keyes 2007)[2]. In any case, it would appear that the term *respeaking* is rapidly consolidating, both in the industry (Marsh 2006) and in academia (van der Veer 2008), as the most common label to refer to what Eugeni (forthcoming a) defines as:

> (…) a technique thanks to which the respeaker listens to the source text and re-speaks it. The vocal input is processed by a speech recognition software which transcribes it, thus producing real-time subtitles.

Later on in the same article, and having explained more in detail some of the ins and outs of respeaking, Eugeni (forthcoming a) provides a more precise definition:

> (…) respeaking is a reformulation, a translation or a transcription of a text, produced by the respeaker, processed by a speech recognition software and

> being broadcast simultaneously with the production of the original text - in our definition, an audiovisual event be it sport, news or other programmes being broadcasted live and requiring real-time subtitling. According to the end users' needs, the output can be displayed in a variety of different formats and colours.

Both definitions provide, from a different perspective, an initial view of what respeaking entails. Most importantly, they point to a number of key ideas that can be further developed in order to gain more thorough insight into the nature of this discipline. One of these ideas has to do with the nouns chosen by Eugeni to refer to respeaking at the beginning of the second definition (*reformulation, translation* and *transcription*). Although all three can be said to apply, the current professional practice of respeaking seems to indicate that it is the first one, reformulation, which describes the most common case-scenario (Marsh 2004). Thus, despite the fact that respeaking can indeed entail translation, understood as an interlingual phenomenon, this only seems to occur in exceptional cases such as that of BBC Wales, where live programmes are respoken from Welsh into English. More often than not, though, respeaking is an intralingual phenomenon aiming at producing live[3] subtitles for the Deaf and Hard-of-Hearing audience within the same country or language community. As for the idea of respeaking as a verbatim transcription of the source text (ST), it opens up a can of worms whose scope would merit an article in itself. Suffice it to note here that a verbatim rendering of a given ST may pose considerable problems not only for the audience but also for respeakers. Indeed, the speech rate of some TV presenters, as well as the need for respeakers to dictate punctuation and change the colour and position of the subtitles to identify speakers and clear the screen, makes completely verbatim respeaking not only "very difficult to follow" (Eugeni forthcoming b) for the audience, but also very difficult to achieve for respeakers.

A second idea that may be derived and developed from the above definitions of respeaking has to do with the technology required. From the point of view of the respeaker (thus excluding the technology used by TV channels to actually broadcast a respoken programme), this consists of speech recognition software integrated into subtitling software. The latter may allow the respeaker to change the subtitle colour and position when needed but it is in essence an interface designed to display the recognised utterances as subtitles on the screen. In this sense, it is the speech recognition software that does most of the work and also the one that requires most work on the part of the respeaker. Indeed, a key part of the respeaking job, and thus of the training of respeakers, is the training of the software, which becomes an indispensable co-worker for the respeaker (Remael and van der Veer 2006). Not only is it a tool, but a partner which, if no corrections are made, is going to have the final say about the

subtitle that will be displayed on the screen. Thus, in the same way that speech recognition software is often described as speaker-dependent (a respeaker is needed as an intermediate step between the ST speaker and the software), the respeaker can be said to be software-dependent.

Finally, there is one more aspect that is particularly worth-noting regarding Eugeni's definitions. As has already been mentioned, they adopt different perspectives to define the same phenomenon. The first one focuses on respeaking as a process whereas the second is concerned with respeaking as a product. As a process, respeaking is somewhat akin to interpreting, in that respeakers are expected to provide a more or less simultaneous translation (in this case intralinguistic), thus splitting their attention to speak the target text (TT) as they listen to the ST. As a product, respeaking may be regarded as the production of non-synchronous subtitles (there is usually a 3-4 second delay) which are usually expected to reformulate or transcribe what is being said by the speaker/s. Far from being a merely theoretical ivory-tower discussion, this distinction between process and product may sometimes be very relevant, as pointed out by Orero (2004) in the case of voice-over. In respeaking, it lies at the root of the current professional consideration of this new discipline. Indeed, drawing on her experience as a respeaker at Red Bee, Marsh (2004:26) explains that:

> Respeaking is not yet recognised as a job in its own right, but merely as a branch of subtitling. Respeakers and subtitlers are in the same salary bracket, even though their jobs entail very different things; remuneration for respeaking, however, is significantly lower than for stenography, even though the two jobs are very similar.

It thus follows that, whereas respeakers are regularly carrying out the process of respeaking (thus listening to the ST and transcribing/ reformulating it as they dictate punctuation and change the subtitle colour and position), they only achieve recognition for what they produce (subtitles displayed on the screen). In other words, their actual job may entail the process, but they are only credited with making the product (Romero Fresco 2008). In this sense, if more attention was paid to the former, their profile would logically have to resemble more that of an interpreter, given that, except for the change of language, similar conditions apply (prior research, stress, cognitive load etc.). In this sense, the present article may be regarded as a small contribution to raise awareness about the actual nature of a discipline that is still to find its rightful place both in academia and in the professional market. With this purpose in mind, the next section is devoted to positioning respeaking in relation to simultaneous interpreting and subtitling.

3. Respeaking, simultaneous interpreting and subtitling: practices in contact?

As stated in the introduction, the object of this study is not merely to compare and contrast respeaking, simultaneous interpreting and subtitling, but rather to determine where the practices approach each other and, in doing so, identify the skills professional respeakers should master.

3.1. Respeaking and simultaneous interpreting

Having already defined respeaking, the following description summarises the practice of simultaneous interpreting:

> In simultaneous mode, the interpreter sits in a booth with a clear view of the meeting room and the speaker and listens to and simultaneously interprets the speech into a target language. (AIIC webpage)

Thus, it would appear that the basic similarity between the two practices is the simultaneous quality of the actions involved: listening and speaking at the same time. Simultaneous interpreters' and respeakers' verbal agility and speed must be activated immediately upon receiving the message. The fact that the two activities share this trait has led many authors to draw a parallel between them[4]. Furthermore, the two activities also share clear time constraints, namely real-time production, little or no margin for correction or improvement and the need for the practitioners to control their voice while listening. Yet, it should be noted here that, whereas interpreters must have good diction, timbre and articulation so that listening is pleasant and comfortable for the audience, respeakers' voices are expected to be flat and monotonous, as they are not addressing a human audience but software that is to recognise their message. Another common feature is the thematic and lexical preparation, achieved in both disciplines through extensive glossaries, databases, and terminology searches. Finally, yet another similarity lies in the working set-up: simultaneous interpreters and respeakers alike work in booths with a microphone and headset. In both cases, it is team work; both team members work for a maximum of 30 minutes at a time alternating work and rest shifts.

As for the differences, an important one is that in simultaneous interpreting the channel the interpreter uses to deliver the TT is the acoustic channel. In respeaking, however, the respeaker listens to a ST, which is oral, and orally produces an intermediate text, but the final TT will be written in the form of subtitles. In addition, as stated by Marsh (2004), the type of work a simultaneous interpreter and a respeaker cover is very different. The former deals with meetings, conferences, summits and court cases, whereas the latter deals with

live television broadcasts such as news, parliamentary sessions, sport events and concerts. As the nature of the work differs, so does the audience.

Yet, the main difference between interpreting and respeaking lies in the fact that the former is always interlingual, and requires decoding in the source language and simultaneous recoding in the target language, whereas respeaking is usually an intralingual activity. This does not make it, as has been mentioned, a repetition of the ST, given that respeaking requires the introduction of pauses and punctuation in speech, as well as a careful selection of the terminology that the speech recognition software can best process. Additionally, extra-linguistic aspects have to be dealt with, for example, by selecting different colours or fonts or making use of labels indicating a change of speaker or a given noise. All of this means that the respeaker, rather than repeating automatically, has to perform a process of message comprehension and reformulation that often requires a certain distancing from mere word-for-word formulation.

This is probably what separates respeaking from shadowing, a technique that has long been used in the initial phases of simultaneous interpreting training. Shadowing consists in simultaneously repeating a speech in the same language, using the same words. Notwithstanding its tradition, several authors have questioned the usefulness of shadowing (Kurz 1992) or rejected it outright (Seleskovitch and Lederer 1989). Their reasoning is that simultaneous interpreting consists of deverbalizing the original and that this exercise leads students to focus too much on the words and not on the idea being conveyed. In his studies on the levels of information processing, Lambert (1993) showed that comprehension and recall are significantly higher when there is a decoding and recoding process than they are in shadowing, which is merely a literal repetition of the source message.

Other authors (Schweda Nicholson 1990; Lambert 1992) argue that shadowing helps students in their initial phases to master the technique of listening and speaking at the same time, in addition to following a pace set by an external source. In this light, shadowing should not be considered as the purpose and product of respeaking, but rather as an exercise that, as in interpreter training, can help future respeakers grapple with the difficult task of listening and speaking at the same time. As in simultaneous interpreting training, it would also be recommendable to introduce, as early as possible in respeaking training, exercises that help the student to avoid following the speaker blindly. Such exercises should help students keep enough distance to understand and reformulate the meaning of the unit to be respoken.

Finally, having explained this, it is important to delve further into the above-mentioned multi-tasking nature of both interpreting and respeaking. Gerver (1971) describes the complex mental activity performed by interpreters as follows:

> (...) interpreters receive and understand a unit of meaning, and begin to mentally translate it and verbally formulate it. At the same time, they receive and understand a new unit of meaning while still occupied in the vocalization of the previous one. Thus, they must be able to retain the second unit in their memory before beginning the interpretation; while they formulate the second unit, they receive the third unit, and so on successively.

Although the presence of the process of translation differentiates interpreting and respeaking as practices, the concept of multi-tasking, as Gerver defines it, would apply to both.

On another note, turning to Moser Mercer's description of the simultaneous interpreting process (1978: 358):

> (...) during the phase of understanding the incoming message, the interpreter connects words with certain conceptual constructions that exist, or are coming into existence, in his memory. In an interpreter these connections are assumed to be of a dual nature: intralingual links (between concept and word in one language) and interlingual links (between the language–specific nodes of the same concept). Given the explicit task of translation, what the interpreter then does is to activate the conceptual relations and arrive at a certain conceptual arrangement, together with activating the necessary intralingual links and expressing this arrangement with target language labels.

Clearly, respeakers, like simultaneous interpreters, must establish conceptual relationships, even if only at an intralingual level. Respeakers perform an information processing that requires a significant cognitive effort, distant from mechanical and automatic repetition, which will have some very direct consequences on the contents and the practice of the training offered to them.

3.2. Respeaking and subtitling

Although a definition of subtitling is probably not necessary, it is important to ascertain what type of subtitling is to be compared with respeaking. In this sense, it may be more logical to draw a comparison with intralingual subtitling for the Deaf and Hard-of-Hearing, given their many similarities. Indeed, they both aim at creating the same product (comprehensible written subtitles in the ST language) for the same audience (mainly Deaf and Hard-of-Hearing viewers, but also hearing viewers who may use them for language-learning purposes or in situations where no sound is heard on screen, such as a waiting room or a pub). For this purpose, respeakers and

subtitlers usually have to reformulate or edit in their own language, often applying text reduction strategies. Good grammar and spelling skills are needed for both disciplines, with particular focus on punctuation, which will have to be delivered orally in the case of the respeaker. Besides, the ST often poses the same type of difficulties for respeakers and subtitlers, namely multiple turn-taking, overlapping dialogue, use of realia (famous names, geographical references, names and institutions) etc. Regarding the audience, both respeakers and subtitlers need to be aware of their viewers' needs and requirements, so as to, for instance, produce appropriate extralinguistic information.

As far as the differences are concerned, the most important are two: the translation situation (offline / live) and the translation mode (written / oral). As regards the former, whereas subtitlers may have more or less time to produce their work, respeakers have to deal with the pressure inherent in a live situation. Thus, all the difficulties involved in the subtitling process that may be shared by subtitling and respeaking, such as the need to deal with technological vagaries, become increasingly demanding in respeaking, where pause, re-thinking and correction are usually not an option. As for the translation mode, it is often very different: whereas subtitlers (when they have access to the written script of the ST) provide a written-to-written translation, respeakers (considering respeaking as a process) provide an oral-to-oral translation. In the light of this, it would appear that, in many ways, respeaking is to subtitling what interpreting is to translation, i.e. a leap from the written to the oral without the safety net provided by time.

4. Competencies and skills of a respeaker, a taxonomy

After comparing and contrasting respeaking, simultaneous interpreting and subtitling, the next step is to identify the competencies inherent in each of the practices in question that respeakers must master to perform their professional duties. Needless to say, the identification of these skills is a fundamental step for the design of any respeaking course.

Thus, the following taxonomy outlines the most relevant skills required for a professional respeaker and arranges them in a double-entry matrix. The vertical columns delimit the fields from where the competences are obtained, be it subtitling, simultaneous interpreting or respeaking, in the case of those that are inherent to this discipline. The horizontal rows feature other elements of classification. The first of them is temporal, and is related to the process carried out by the respeaker. A distinction is made here between the skills to be activated before the process and those to be applied as the process is taking place. In turn, the latter are further divided into those skills

that are related to the ST, the TT or the transition between the two, which is referred to as crossover.

RESPEAKING SKILLS			
	FROM SDH	**FROM SIMULTANEOUS INTERPRETING**	**SPECIFIC TO RESPEAKING**
PRIOR TO THE PROCESS	<u>Software-related skills</u> -Ability to use industry-standard subtitling software.	<u>Preparation skills</u> - Fostering of research capacity; - Ability to develop subject matter glossaries and databases; - Familiarity with terminology in specialized fields; - Compliance with the Code of Ethics. <u>Strategic skills</u> -Ability to work as part of a team.	<u>Software-related skills</u> • General knowledge - Understanding of how it fits in the bigger picture; - Understanding of how ASR technology works (bottom-up as opposed to top-down); - Comfort dictating to ASR software; - Awareness of software demands and limitations; - Confidence in new technology (working with it, not against it). • Preparation of the software - Ability to search for a controlled outcome, anticipating possible errors; - Ability to constantly improve both voice model and vocabulary; - Mastery of the skills required to develop macros; - Readiness to test and report benefits and drawbacks.

DURING **THE PROCESS**	**ST**	<u>General ST-related skills</u> - Ability to reformulate/edit; - Ability to apply text reduction strategies; - Spotting skills. <u>Specific ST-related skills</u> - Ability to cope with multiple turn-taking or overlapping dialogue; - Attention to realia; - Ability to adapt to different programme genres.	<u>Listening comprehension skills</u> - Attention to concentrated listening; - Familiarity with accent recognition: geographic and social variants; - Familiarity with the cultural context; - Ability to develop short term memory; - Ability to develop emergency strategies when ST is not understood **Analysis, synthesis and reformulation skills** - Ability to understand the communicative intention of the source message; - Ability to understand the red thread of the discourse; - Capacity to select and focalize the relevant information; - Ability to divide between main and secondary ideas; - Capacity to identify the discourse connectors; - Capacity to deduce meaning through context and extralinguistic elements; - Ability to condense information; - Ability to segment information in sense units.	

	CROSSOVER	Synchronisation skills - Ability to synchronise oral dialogue and subtitles.	Multitasking skills - Ability to speak while listening; - Capacity to receive analytically the incoming message; - Capacity to monitor the outgoing message; - Ability to keep up with rapid dialogue; - Confidence in maintaining décalage. Live skills - Ability to show calmness and accuracy under pressure; - Capacity to manage stress; - Ability to correct one's own mistakes; - Capacity to cope with frustration caused by the inevitability of mistakes: no second chances; - Attentiveness to keep TT audience in mind.	Multitasking skills - Ability to cope with up to four types of simultaneous intersemiotic tasks: Listening while speaking, writing and reading. Live skills - Ability and speed to change the subtitle colour/label; - Ability and speed to change the subtitle position; - Mental-pre-editing skills; - Ability to cope with visual feedback from screen; - Comfort in dealing with no feedback from the audience; - Confidence in dealing with technological vagaries.
	TT	Production skills - Accurate grammar; - Accurate spelling (including punctuation); - Ability to produce comprehensible subtitles. Awareness of target audience - Understanding of Deaf people's difficulties watching TV; - Knowledge about the required reading speed (intuitive sense of reading speed for segmentation); - Ability to convey extra-linguistic information; - Awareness of style sheets and norms.	Delivery skills - Ability to express thoughts clearly and concisely; - Extensive vocabulary; - Ability to control one's own voice; - Ability to communicate fluently; - Ability to communicate with accuracy; - Capacity to reproduce the same tone and register as in the ST; - Capacity to transmit conviction and self confidence; - Ability to communicate with good voice projection without hesitations, unnecessary repetitions and corrections; - Ability to communicate with good diction.	Delivery skills - Accurate oral punctuation; - Ability to speak in short stretches of text; - Consistency of delivery; - Ability to produce a flat and clear pronunciation; - Ability to set word boundaries clearly; - Attentiveness to enunciate sounds within small words; - Attentiveness to avoid pauses between each word; - Ability to dictate at higher than average speed.

Table 1: Taxonomy of Respeaking Skills

Whereas most of the skills included in the above table are self-explanatory or have been described in section 3, many of those classified as 'specific to respeaking' deserve further explanation.

First of all, before actual respeaking can be carried out, respeakers must be fully familiarised with the software they are going to use, especially with their Automatic Speech Recognition (ASR) software. Respeakers are expected to know, for example, how ASR technology fits in the bigger respeaking picture, that is, how much of the end-result hinges upon the performance of this software. This is very important, as the amount of work carried out by this software in the respeaking process is directly proportional to the amount of work that must be carried out by the respeaker to constantly train it and improve it beforehand. Likewise, respeakers are expected to have a basic understanding of how ASR technology processes acoustic data, which is very different to the way humans do. As Keyes (2007) points out, we adopt a top-down approach to recognise speech, thus resorting to concepts and circumstantial knowledge to distinguish words. Computers adopt a bottom-up approach, analysing sound structures, the most basic of which is the phoneme, to be able to recognise speech. This is of paramount importance for the delivery of the TT, as respeakers are expected to pronounce words carefully, setting clear boundaries between them to minimise misrecognitions. Overall, respeakers must feel at ease when dictating to the ASR software, which requires familiarisation with the software demands and limitations. Once these limitations are identified, respeakers must try to either overcome them or at least minimise them through training, which is possibly the most important part of the preparation stage in respeaking.

In general, the main aim of this training stage is to obtain a conflict-free outcome (Keyes 2007), that is, to make sure that the ASR software departs as little as possible from what the respeaker has dictated. At least three tools are available for this purpose: individual voice models, vocabularies and macros (Eugeni forthcoming a). First of all, respeakers have their own voice models, which they create and enhance through continuous dictation, thus helping the software to 'get used to' their speech patterns. Although current ASR software such as the English version of Viavoice, used by Red Bee, has a sizeable corpus of some 100,000 words, many specialised terms or proper nouns needed to respeak a specific (audiovisual) programme are likely to be missing. Respeakers must introduce them through dictation, thus fine-tuning their voice models and minimising the error rate. Yet, as pointed out by Marsh (2004), the software is bound to have problems when deciding between homophones or near-homophones such as *bunker* or *banker*. In this case, respeakers can make use of a second tool – the creation of specific vocabularies

for specific topics. Thus, for the golf vocabulary they will introduce most of the specific terms used in this sport and, once it is activated, the ASR software will choose *bunker* instead of *banker*, in the same way that it would opt for the latter should the financial vocabulary be selected. Finally, also very useful are the so-called macros. In this case, respeakers can set the software to display a word or group of words every time they utter a given command, which they can make up. This can be helpful to save much-needed time when respeaking (for instance, the command *Queen macro* could trigger 'Her Majesty the Queen Elizabeth II'), but also to avoid potential misrecognitions (*Bor-macro* to trigger the surname Borowski as opposed to, say, 'brought ski'), to improve punctuation (*mac-ex* for exclamation mark) and to change the subtitle colour orally (*macroyellow*). It thus follows that a great deal of the preparation work to be carried out by respeakers beforehand lies in being able to anticipate the potential problems that may be faced by the software and in using the available tools as effectively as possible to solve them.

As for the respeaking skills included within the 'crossover' category, one that is worth noting is mental pre-editing (Remael and van der Veer 2007). Indeed, apart from the multitasking process involved in listening (ST) while speaking (TT), typing (subtitle position and colour) and reading (the subtitles, trying to pay attention to the potential errors), respeakers are also expected to anticipate what will or will not be recognised by the software. Many terms may come up that have not been prepared beforehand, and so respeakers will have to find a way round them to avoid potential misrecognitions. Finally, the delivery skills needed for respeaking are fairly self-explanatory, but it should be pointed out that they may also depend on the ASR software used or, more accurately, on the way it displays the subtitles on the screen. Viavoice, for example, has a word-for-word display mode, whereas Dragon or Vista display subtitles in chunks which usually correspond to full sentences. There is in this case a bigger delay, as the software waits for the respeaker to dictate the last word of a chunk/sentence to show the whole utterance on the screen. Thus, when respeaking with Viavoice, respeakers are basically concerned with being able to split their attention to listen as they speak, type, and read. With Dragon or Vista, respeakers may be expected to minimise the inevitable delay by producing short sentences that will be shown as subtitles, and thus need to bear in mind their appropriate length and other relevant features, which means that they need “insight into subtitling concepts such as reading speed and spotting” (Remael and van deer Veer 2006). Therefore, although all the skills included in the table above are necessary, respeaking with Viavoice requires mainly those that are common to interpreting, whereas respeaking with Dragon or Vista also draws heavily on subtitling skills.

5. Training

This section offers a description of different exercises that can be carried out in the classroom so as to provide students with the most important skills outlined above.

5.1. Preparation

Skill to be acquired: ability to train the software for an efficient performance.

Students will be asked to prepare word lists with all the relevant terminology, including both technical terms and recurring topical items of a more general nature. Different types of tests could be run to ensure that they are familiar with the topic. Students will then dictate the word lists to the software, deciding in every case whether the terms should be included in their general voice model or in specific vocabularies. During dictation, students will be asked to pay special attention to the pronunciation of unfamiliar names and other proper nouns, given that they will not be recognised by the software during the respeaking process unless they are pronounced exactly how they were dictated the first time round. For this purpose, an indication of their pronunciation may be added to the word list.

At least two different exercises may be run to assess whether the training of the software is carried out satisfactorily. In the first one, students are given the whole transcription of the ST to be respoken. They read it carefully in order to anticipate all the potential recognition errors that may arise (with particular attention to homophones) and train the software by dictating word lists, updating their voice model and the relevant vocabulary as well as setting the necessary macros. Then, they respeak the text and ascertain whether the errors that have occurred could have been avoided with further training. A more difficult variation of this exercise would consist in giving students a headline of a news item, on the basis of which they have to train the software with the right terms, macros etc. Once again, they would respeak the text and check the results against the training carried out previously.

5.2. Analysis and listening comprehension capacity

Skill to be acquired: ability to analyse the ST, detecting the sense units and reducing them to be processed in the memory storage (segmentation).

For this purpose, an activity could consist in having students listen to a speech, without taking notes, and then answer a number of

questions related to the content presented. Another exercise could be based on orally summarizing a speech they have just heard. Yet another drill could be built on the introduction of deliberate difficulties in speeches so that students are able to employ coping tactics. These difficulties may range from fast speech rates to unplanned structures, redundancies, speaker hesitation and/or missing links. Finally, anticipation skills could be strengthened by reading speeches and leaving out the end, and then having students follow their logic to anticipate the content with which the discourse could continue.

5.3. Synthesis and reformulation capacity

Skill to be acquired: ability to identify the key elements of the ST, discard superfluous items and apply reformulation/editing strategies.

First of all, students may be asked to listen to a speech and then a) identify main ideas; b) establish a list of key words and links; c) create a conceptual map of ideas: hierarchy and relationship among them. As for reformulation exercises, students could listen to an oral text and then repeat it in their own words, first sticking to the text, then simplifying it and finally presenting a more flowery version or changing the register (colloquial/ formal). Along these lines, Gillies (2001) proposes an exercise that involves inverting the meaning of the text. Following the same idea, he proposes that students rework the grammatical structure of sentences without changing their meaning, i.e., change all passive verbs to indicative, remove subordinate clauses, etc. Students are also to change the stance adopted by the speaker in a speech or to make it more or less serious or ironic. Another interesting exercise consists of practising changing the order of the clauses in a sentence without changing its meaning.

As suggested by Remael and van der Veer (2006) and in order to hone their subtitling skills, students may be given the transcription of a ST, which they have to segment and rewrite as subtitles that are appropriate for the intended audience. Once this has been accomplished, they could do the same exercise listening to the ST, as opposed to reading its transcription, and stopping it every short while to write/type the subtitles. This would help not only their segmentation skills but also their short-term memory.

5.4. Multitasking capacity

Skills to be acquired: ability to speak while listening; confidence in maintaining *décalage*.

There are several introductory exercises that help students experience the feeling of splitting their attention, such as listening to a narration while counting to a hundred, or reciting a well-known

poem while listening to a speech and then summing up the speech before an audience. Yet, at this stage it is important for students to get used to dictating punctuation marks as they speak. This could be done with one of the most controversial exercises in this realm – shadowing. Although, as explained in section 3, this exercise is often criticised for making students follow the ST words literally disregarding the content, the correct introduction of punctuation marks already requires a considerable degree of processing that cannot be done with a parrot-like repetition of the ST. Furthermore, to prevent students from following speakers too closely and ensure sufficient décalage to understand the meaning of the unit to be reformulated, there is an exercise that Van Dam (1989:170) calls the *distance exercise*. It is made up of two phases. In phase one, the teacher pauses following every idea for the student to reformulate it. The second phase consists of beginning the enunciation of the next idea while the student is still in the midst of reformulating the prior one.

5.5. Dealing with a live situation

Skill to be acquired: ability to overcome the stress caused by a live situation.

A series of improvisation exercises could be carried out to foster the students' self-confidence when respeaking. For example, a student may improvise a 3-minute speech on a subject volunteered by a colleague. Other students listen and comment on the speech. This exercise trains the delivery technique as well as the split attention of the students since, as they improvise, they must be thinking ahead about the next sentence. Another option would be to have one student rendering a speech, while students outside the booth show cards with keywords at short intervals. The student giving the speech must incorporate the word or idea coherently into the improvised discourse.

Alternatively, Gillies (2001) advises students to practice in the most relaxed position they can come up with. This should counterbalance the unnaturally tense posture of most students.

5.6. Delivery

Skills to be acquired: ability to express thoughts clearly and concisely, transmitting conviction and self-confidence; ability to dictate in short stretches of text at higher than average speed; ability to dictate with a flat and clear pronunciation, including oral punctuation, setting boundaries between words and anticipating potential software errors.

The improvisation exercises described in the previous section could be a good way to train delivery techniques. Although in their preparation of the software, students will already have had the opportunity to practise dictation, it is important that they get used to doing so at a higher than average speed. For this purpose, they may be asked, first of all, to respeak a written script and then assess their speed (words per minute) and accuracy (error rate), which is expected to range from 95% to 97%. Then, they may be asked to do the same exercise at a higher speed, aiming at a target of 120-140 words per minute. As is often the case in voice writing competitions (Vincent 2007), this could be done by having a clock running on the computer screen as the student respeaks the ST. After every paragraph there will be an indication of the time it should have taken the respeaker to dictate until that point if s/he is to meet the target set at the outset. Thus, not only will this foster the students' speed and accuracy when dictating but also their multitasking skills.

Once students have achieved the required accuracy and speed in dictation, they can practise respeaking proper. First of all, they could respeak the STs for which they have already prepared the subtitles, both from the written transcription and the audio version. Then, they may move on to a more real-life scenario listening to a programme several times and respeaking it, and finally respeaking a programme as they listen to it for the first time. The preparation exercises outlined above (such as training the software to respeak a programme on the basis of just one headline) may be incorporated here.

Furthermore, other difficulties may be added in order to strengthen the students' ability to anticipate and solve potential recognition errors. The teacher could have students respeak a text for which they have prepared their software but add new terms that have not been prepared beforehand. Students will thus have to find a way round these terms to avoid potential misrecognitions. An alternative exercise would be to prevent them from using certain terms that will recur in the text. In addition, so as to ensure that the respoken subtitles are not too long (given the audience they are intended for), students may also be asked to stick to, for example, one liners, which means that they will also have to split their attention, checking the screen as they respeak the ST.

Finally, an essential part of the training of respeakers, and one that may apply to most of the exercises mentioned here, is the type of materials used. If the aim is, as advised by Remael and van der Veer (2006), to use industry-standard material, this could consist of sports (such as golf, football or snooker), public events (such as opening speeches), parliamentary sessions and news (whether news broadcasts or debates). Following the training provided at Red Bee

(Marsh 2006), sports could be used at an initial stage, given that only certain utterances are expected to be respoken, namely those that add something to what is already being seen on the screen. Public speeches could constitute the next stage, as they require more constant respeaking but there is usually no need to change subtitling position or colour for different speakers. Once these two genres are mastered, students may move on to respeaking different types of news (with speech rates up to 180 words per minute) and finally to the more demanding task of respeaking parliamentary sessions or lively debates (high speech rate, quick turn-taking, overlapping etc.).

6. Final remarks

Slowly but surely, respeaking seems to be consolidating as the preferred method to provide intralingual live subtitles for the Deaf and Hard-of-Hearing in many TV channels. Yet, as regards research and especially teaching, respeaking is far from being consolidated. Apart from the one-off course carried out at Università Di Bologna (Forlì), only Universiteit Antwerpen offers consistent training in respeaking, with a semester-long module included in its Masters in Interpreting. Other universities are beginning to devote some hours to respeaking as part of the intralingual subtitling modules in their Masters on AVT (Universitat Autònoma de Barcelona, Università Di Bologna), but respeaking courses are in general few and far between, which explains why TV channels are still choosing to train their own respeakers from scratch.

In this sense, there seems to be a certain lack of confidence in respeaking as a subtitling technique at university level. This may be due to the widespread and somewhat hasty release of ASR software in the mid-to-late 90s, with very average results (Theriod 2007). The memory of this, added to the average results usually obtained in respeaking workshops at AVT conferences (which is inevitable, given that there is no time to train the software), have resulted in reservations as to the reliability of respeaking and in the consideration of this technique as “a glimpse of the future” (Lambourne 2007). Needless to say, the current professional practice of respeaking in many TV channels shows that this is very much a present reality, and one that could open up new job opportunities for both subtitlers and interpreters.

With a view to training students to become professional respeakers, the present article argues for specific training in this area, based on the skills required (whether from interpreting, subtitling or specific to respeaking) and the exercises with which they may be acquired. Yet, this is merely a proposal, and thus more research is needed not only at this level, but also regarding related areas such as the materials that should be used in a respeaking course or the stages this course

should consist of. Likewise, further thought should be given to other aspects such as the trainers of this discipline (their qualifications, their training), the status of a potential respeaking course at university (graduate or postgraduate, as part of an interpreting course, an AVT course, independent) and many other questions that demand immediate answers for a present reality.

References

- **AIIC web page**. On line at: http://www.aiic.net (consulted 26.02.2008).
- **Baaring, Inge** (2006). "Respeaking-based online subtitling in Denmark". *Intralinea, Special Issue*. http://www.intralinea.it/specials/respeaking/eng_open.php (consulted 26.2.2008).

- **Eugeni, Carlo** (2006). "Introduzione al rispeakeraggio televisivo". http://www.intralinea.it/specials/respeaking/eng_more.php?id=444_0_41_0_M (consulted 26.02.2008).

- **Eugeni, Carlo** (2008). "A Sociolinguistic Approach to Real-time Subtitling: Respeaking vs. Shadowing and Simultaneous Interpreting". C.J. Kellett Bidoli and E. Ochse (Eds), English in International Deaf Communication, Linguistic Insights series vol. 72, Bern: Peter Lang, 357-382.

- **Eugeni, Carlo** (forthcoming, a). "Respeaking at BBC". *Intralinea, Special Issue*.

- **Eugeni, Carlo** (forthcoming, b). "Respeaking a political debate for the Deaf: the Italian case".

- **Gerver, David** (1971). *Aspects of Simultaneous Interpretation and Human Information Processing*. M.A.Thesis. Oxford University. Gillies, Andrew (2001). *Conference Interpreting – A Student´s Companion*. Cracow: Tertium.

- **Kalina, Sylvia** (2000). "Interpreting Competences as a Basis and a Goal for Teaching". *The Interpreters' Newsletter*,10, 3-32.

- **Keyes, Bettye** (2007). "Realtime by Voice: Just what you need to know". Paper presented at Intersteno Congress in Prague in July 2007.

- **Kurz, Ingrid** (1992). ""Shadowing" Exercises in Interpreter Training". Dollerup, C. and Loddegaard, A. (Eds.). *Teaching Translation and Interpreting. Training, Talent and Experience*. Amsterdam/Filadelfia: John Benjamins Publishing Company, 245-250.

- **Lambert, Sylvie** (1992). "Shadowing". *Meta*, 37 (2), 263-273.

- **Lambert, Sylvie** (1993). "The effect of ear of information reception on the proficiency of simultaneous interpretation". *The Interpreters Newsletter*, 5, 22-34.

- **Lambourne Andrew, Jill Hewitt, Caroline Lyon & Sandra Warren** (2004) "Speech-Based Real-Time Subtitling Services". *International Journal of Speech Technology* 7 (4), 269–279.

- **Lambourne, Andrew** (2007) "Real-time Subtitling: Extreme Audiovisual Translation". Paper presented at the conference *LSP Translation Scenarios* in Vienna in May 2007.

- **Marsh, Alison** (2004) *Simultaneous interpreting and respeaking: a comparison.* Unpublished MA Thesis. University of Westminster.

- **Marsh, Alison** (2006) "Respeaking for the BBC". *Intralinea, Special Issue*.http://www.intralinea.it/specials/respeaking/eng_more.php?id=484_0_41_0_M (consulted 27.02.2008).

- **Moser Mercer, Barbara** (1978) "Simultaneous Interpretation: A Hypothetical Model and its Practical Application". D. Gerver & H. Sinaiko (Eds.), *Language, Interpretation and Communication.* New York: Plenum Press, 353-368.

- **Muzii, Luigi** (2006) "Respeaking e localizzazione". *Intralinea, Special Issue*. http://www.intralinea.it/specials/respeaking/eng_open.php (consulted 26.02.2008).

- **Orero, Pilar** (2004) "The Pretended Easiness of Voice-over Translation of TV Interviews". *The Journal of Specialised Translation*, 02, 76-96. http://www.jostrans.org/issue02/art_orero.php. (consulted 31.03. 2008).

- **Orero, Pilar** (2006) "Real-time subtitling in Spain". *Intralinea, Special Issue*. http://www.intralinea.it/specials/respeaking/eng_more.php?id=450_0_41_0_M (consulted 26.02.2008).

- **Padilla, Presentación & Bajo, Teresa** (1998). "Hacia un modelo de memoria y atención en interpretación simultánea". *Quaderns. Revista de traducció*, 2, 107-117.

- **Remael, Aline & Bart van der Veer** (2006). "Real-time Subtitling in Flanders: Needs and Teaching". *Intralinea, Special Issue*. http://www.intralinea.it/specials/respeaking/eng_open.php (consulted 26.02.2008).

- **Remael, Aline & Bart van der Veer** (2007). "Teaching live-subtitling with speech recognition technology: what are the challenges?" Paper presented at the conference *LSP Translation Scenarios* in Vienna in May 2007.

- **Romero Fresco, Pablo** (forthcoming, 2008). "La subtitulación rehablada: palabras que no se lleva el viento". Pérez-Ugena, Álvaro & Vizcaíno-Laorga, Ricardo. *ULISES: Hacia el desarrollo de tecnologías comunicativas para la igualdad de oportunidades,* Madrid: Observatorio de las Realidades Sociales y de la Comunicación.

- **Schweda Nicholson, Nancy** (1990). "The Role of Shadowing in Interpreter Training". *The Interpreters' Newsletter*, 3, 33-37.

- **Seleskovitch, Danika & Lederer, Miriam** (1989). *Pédagogie raisonnée de l´interprétation*. Bruxelles-Luxembourg: Didier érudition Opoce.

- **Theriod, Chad** (2007). "Speech Recognition as a Rich media Component". Paper presented at Intersteno Congress in Prague in July 2007.

- **Van Dam, Ine-Marie** (1989). "Strategies of Simultaneous Interpretation". Gran, Laura & Dodds, John. *The Theoretical and Practical Aspects of Teaching Conference Interpretation*, Udine: Campanotto Editore, 167-176.

- **Van der Veer, Bart** (2007). "De tolk als respeaker: een kwestie van training". *Linguistica Antverpiensia*, LA NS6, 315-328.

- **Vincent, Keith** (2007). "A Brief presentation to Intersteno participants", paper presented at Intersteno Congress in Prague in July 2007.

Biographies

Dr Marta Arumí Ribas holds a degree in Translation by the Universitat Autònoma de Barcelona (UAB), a PhD in Interpreting by the Universitat Pompeu Fabra (UPF), a Master´s Degree in Conference Interpreting by the Universidad de La Laguna (ULL) and a Master´s Degree in Teaching Conference Interpreting by the École de Traduction et Interprétation of the Université de Genève. She lectures at the UAB, where she coordinates a Master Degree in Conference Interpreting. Her main lines of research are the teaching of consecutive interpreting and self-regulation processes in the classroom. She participates in various funded research projects and groups. She has been working as free-lance conference interpreter for more than ten years.

Marta.Arumi@uab.cat

Pablo Romero Fresco completed his PhD on the naturalness of the Spanish dubbing language at Heriot-Watt University (UK), where he taught Translation, Liaison Interpreting, Subtitling and Respeaking. He is now a Lecturer in Audiovisual Translation at Roehampton University, where he teaches Dubbing, Subtitling and Respeaking. He also teaches Respeaking at the MA on Audiovisual Translation at Universidad Autònoma de Barcelona and at the Università di Bologna (Forlì), as well as a number of tutorials for the MA on Translation Studies at University of Edinburgh. He is a member of the research group Transmedia Catalonia, for which he is working on D'Artagnan, a European research project exploring the possibility of providing a common standard for Subtitling for the Deaf and Hard of Hearing.

P.Romero@hw.ac.uk

_____

1. This article has been written in the framework of the research project “La subtitulación para sordos y la audiodescripción: primeras aproximaciones científicas y su aplicación” (HUM2006-03653FILO), funded by the Spanish Ministry of Education.

2. *(Real-time) voice-writing* is a common term in USA to refer to the use of speech recognition to produce not only live subtitles but also transcriptions in trials, classes and different types of public events.

3. It should be noted that, at least in Red Bee, company providing subtitles for the BBC, respeakers also make off-line subtitles, which are then checked by pre-recorded subtitlers (Marsh 2006). This is called *scripting*.

4. An exhaustive comparison between both practices can be found in Eugeni (2008).