Feasibility, Quality and Assessment of Interlingual Live Subtitling: A Pilot Study

🆔 **Hayley Dawson** ✉
University of Roehampton

_____

Abstract

Intralingual respeaking has been widely practiced since 2001 (Romero-Fresco, 2011); however, interlingual respeaking (from one language into another) is yet to take off. Interlingual respeaking is a hybrid form of subtitling and interpreting and calls upon skills used in both professions. To consolidate this mode of audiovisual translation (AVT) within media accessibility (MA), a programme must be created to train future interlingual respeakers. This paper presents the results of the first ever study on interlingual respeaking, in which 10 participants interlingually respoke three short videos using a language combination of English and Spanish. The main areas of research in this project are feasibility, quality and training. Before expanding training in this area, interlingual respeaking must be deemed feasible and an effective method of assessment must be in place to determine its quality. The NTR model is a quality assessment model for interlingual live subtitles, of which an accuracy rate of 98% or above indicates acceptable live subtitles. The average accuracy rate of the study is 97.37%, with the highest accuracy rate reaching the 98% threshold with 98.50%. The initial results point to interlingual respeaking as feasible providing a training programme is put in place to build upon existing task-specific skills and develop new ones to ensure interlingual live subtitles of good quality are produced.

Key words: interlingual live subtitling, quality, NTR model, training, respeaking, media accessibility (MA).

✉ dawsonh@roehampton.ac.uk, https://orcid.org/0000-0001-7156-1233

1. Introduction

Respeakers use speech recognition software to repeat or paraphrase what is heard from an audiovisual text through a microphone while enunciating punctuation and adding special features, such as colours to identify the speakers. The software turns the spoken utterances into text on screen, which are cued as live subtitles (Romero-Fresco, 2011). Although intralingual respeaking has become an established practice within the industry, interlingual respeaking is yet to take off. Intralingual respeaking has been widely practiced since 2001, when it was first used in the UK as a method to provide live subtitles for the BBC World Snooker Championships (Romero-Fresco, 2011). Interlingual respeaking mirrors a similar process: a respeaker listens to an audiovisual text in its original language and respeaks it into another. The respeaker essentially simultaneously interprets what they hear, enunciates punctuation and endeavours to correct errors and add special features for a deaf and hard-of-hearing audience before cueing the subtitles. The shift in language adds a layer of complexity to interlingual respeaking. It is not widely practised and must be researched before better quality live subtitles can be produced.

Many aspects of intralingual live subtitling have been researched, such as sociolinguistic approaches to respeaking (Eugeni, 2008), training (Arumí Ribas & Romero-Fresco, 2008), (Romero-Fresco, 2012); and quality and audience reception (Romero-Fresco, 2012, 2015, 2016; Fresno, 2019). Such research has successfully informed the field of the challenges that lie ahead for intralingual respeaking. Broadcasters have increasingly used the NER model (Romero-Fresco & Martínez, 2015) due to a rise in focus on the quality of live subtitles. The model analyses the extent to which errors affect the coherence of the subtitled text or modify its content. As a next step towards improving quality in intralingual respeaking, respeaking certification has begun with LiRICS (Live Reporters International Certification). LiRICS aims to set and maintain high international standards in the respeaking profession, which is a step in the right direction to highlight respeaking as a qualified profession (Romero-Fresco et al., forthcoming).

Interlingual respeaking is considered a new discipline within the realm of AVT and MA. The ILSA project (Interlingual Live Subtitling for Access) is a large-scale project, which aims to design, develop, test and validate the first training course for interlingual respeaking. The project has become a starting point for research to take place and has explored the complex nature of the task including: research into understanding who live subtitlers are (Robert, Schrijver, & Diels, forthcoming), the task-specific skills required for interlingual respeaking (Pöchhacker & Remael, forthcoming) and what interlingual respeaking performance means for training in interlingual respeaking (Dawson & Romero-Fresco, forthcoming). SMART (Davitti, Sandrelli, & Romero-Fresco, 2018) is a similar pilot study carried out at the beginning of 2018 and aimed to compare the interlingual respeaking performance of trainees with previous experience in interpreting, subtitling and respeaking. The research conducted for this study (in February 2017) and the SMART pilot study (in January 2018) has already shed light on the feasibility and quality of interlingual respeaking and the task-specific skills required. Comparisons have been made with this pilot study

and the SMART study throughout this article. Although there is a small number of participants in this study, it is the first step towards launching and informing the practice of interlingual respeaking. The results have already successfully informed the main experiment of Intellectual Output 2 of the ILSA project.

The service of live subtitles has mainly been provided intralingually and is primarily for those with a hearing impairment but also caters for a hearing audience when sound cannot be used on the television such as in gyms, cafes and libraries. Over the past few years, an interest in interlingual respeaking has grown, which is prospective of revolutionising MA. Interlingual respeaking could heighten social impact as it not only caters for those with a hearing impairment. It also caters for foreign audiences, which demonstrates the potential to aid the integration of foreigners into society. Interlingual respeaking gives those access to a text, which, due to a language barrier, they would not usually be able to access in its original form. This mirrors the idea that MA can be an effective tool to foster human rights for all citizens, not only for those with disabilities (Greco, 2016). Shaping the training of interlingual respeakers is essential in implementing sturdy theory and techniques on how respeakers work in the UK and abroad. This is the key to contributing to the broadened scope of MA. Most importantly, it ensures that a DHOH (Deaf and Hard of Hearing) and foreign community can fully access media products and events in a different form and with live subtitles of good quality.

Training in respeaking has been included in AVT modules in universities across Europe (University of Antwerp, Autonomous University of Barcelona, University of Leeds, University of Roehampton, and Zurich University) and has typically focussed on intralingual respeaking with an introduction to interlingual respeaking included at the end of the module. Training in interlingual respeaking has now begun with the first known course, which was delivered online by the University of Vigo to seven students from January to June 2019. The course included three modules: simultaneous interpreting (English > Spanish), intralingual respeaking (Spanish > Spanish) and interlingual respeaking (English > Spanish); indicating that current professionals and students have started to take interest in the most recent mode of translation and interpreting.

This article, firstly, endeavours to present relevant data outlining the feasibility of interlingual respeaking and how quality has been measured using the current working NTR model (Romero-Fresco & Pöchhacker, 2017) (defined in section 3). Secondly, qualitative results on participants' perception of their performance, the task-specific skills and best-suited professional profile for an interlingual respeaker will be presented. Then, a brief evaluation on the effectiveness of the NTR model will be given. Finally, initial thoughts on training will be presented as a step towards producing interlingual live subtitles of good quality. The findings of this study aim to inform the design of a large-scale study of around 50 participants, which will seek to identify the task-specific skills and best-suited professional profile for an interlingual respeaker. The outcomes will eventually inform a training model for interlingual respeaking.

2. Methodology

The methodology of the pilot experiment took a train, practise and test approach. Participants filled out a pre-experiment questionnaire, a short session on respeaking was delivered, participants completed a respeaking test and filled out a post-experiment questionnaire. Finally, participants contributed further by answering some questions and making observations about their experience of respeaking. Quantitative data was collated in the form of analysis of the respoken subtitles and qualitative data in the form of questionnaires. Individual performance was recorded with Screencast recording software and analysed separately.

2.1.   The Participants

This study took place in the language computer laboratory at the University of Roehampton. Participants received face to face training and the researcher was present to lead the training and experiment. Ten participants took part in the study. However, due to technical issues the data from participant 5 is not available and there is not enough data from participant 10 to carry out a meaningful analysis. Of the remaining eight participants, seven were female and one was male. Their average age was 32, the youngest being 23 and the oldest 48. There were seven native Spanish participants and one native English (participant 7). Two participants were professionals, one being a translation lecturer and another a speech-to-text interpreter. Six participants were postgraduate students in translation, two of whom also worked in translation and teaching while studying.

Table 1.

Details of Participants' Previous Experience

	Background
Participant 1	Subtitling, interpreting and intralingual respeaking
Participant 2	Subtitling and respeaking
Participant 3	Subtitling
Participant 4	Subtitling and interpreting
Participant 6	Subtitling and intralingual respeaking
Participant 7	Subtitling and interpreting
Participant 8	Interpreting and intralingual respeaking
Participant 9	Subtitling

2.2.   Data Collection

Participants respoke three video clips interlingually: a narration, a speech and a news story. Due to two sets of language combinations (English into Spanish and Spanish into English) video clips of similar genres were made available in both Spanish and English. Only two participants respoke the news clip so the results have not been included in the quantitative analysis.

Table 2.

Details of Video Clips Used During the Pilot Experiment

Language combination	Genre	Description	Duration	Words per minute (wpm)
ES > EN	Narration	Wildlife documentary	2 mins 24 secs	73 wpm
ES > EN	Speech	Presidential speech	2 mins 24 secs	131 wpm
ES > EN	News	RTVE - Robot museum	1 min 58 secs	191 wpm
EN > ES	Narration	*Desperate Housewives*, opening scene	2 mins 33 secs	102 wpm
EN > ES	Speech	Presidential speech	2 mins 4 secs	101 wpm
EN > ES	News	BBC - *Can a robot do your job?*	2 mins 9 secs	173 wpm

The video clips used were chosen to represent a variety of content that would usually be respoken in a real-life scenario. The narration and speech videos had one speaker and the news videos had multiple speakers. The narration videos were chosen to allow participants to carry out interlingual respeaking exercises with low speech rates and long pauses. The wildlife documentary clip covered animals which inhabit the Sahara and a clip from the opening scene of *Desperate Housewives* gave insight into a character's life before the show began. Although the genres are different, these clips were chosen as they were both delivered by a narrator and did not include specialist terminology. The two speech and news videos were chosen to replicate content in the form of a live event (speech) and live television (news). The English speech was of former President Barack Obama delivering his farewell speech for the American news channel NBC. The Spanish speech was of President Mariano Rajoy announcing former King Juan Carlos' abdication of the throne. The English news clip covered the public's views on robots carrying out professional jobs and the Spanish news clip covered the opening of a robot museum in Spain. It was hoped that slightly varying speech rates would allow for further analysis of participants' performance when respeaking at different speeds, which has been considered during analysis.

Participants were given the option to watch each video before respeaking it and attempt each one more than once. Only the first attempts were analysed for this pilot study. Participants were given one hour to work on the tasks and respeak each video into their native language through

a microphone attached to the speech recognition software, Dragon NaturallySpeaking. Since no subtitling software compatible with Dragon was available, participants were advised to place a reduced size window of DragonPad underneath the video. DragonPad is similar to a Word document and it allowed participants to make a text box underneath the video to simulate the effect of live subtitles. Screencast software captured each attempt, which recorded the mouse and keyboard movements on screen and audio from the microphone.

Participants completed a pre-experiment questionnaire before any training was given. The questionnaire was composed of the following sections: biographical information, language skills, training, competence, subtitles and respeaking. Closed questions served to determine the demographic of the sample and for participants to rate their own competence in subtitling and interpreting. In the training, competence and subtitles sections, most questions were multiple choice to reflect the limited options for response. The respeaking section was composed of open-ended questions allowing participants to express their current perceptions of respeaking and how they thought they might perform.

After the interlingual respeaking tests, participants completed a post-experiment questionnaire composed of the following sections: level of difficulty, expectations, performance and skills. Level of difficulty and performance required participants to rate their performance and share the most difficult elements of the exercises. Expectations and skills allowed participants to reflect in detail upon what happened during the exercises and note how they perceived their own performance. Participants' perceptions on the skills and best-suited profile for an interlingual respeaker were sought before and after the test.

3. Quality Assessment[1]

Intralingual respeaking has been practised in the UK since the BBC tested it in April 2001 with the World Snooker Championship (Romero-Fresco, 2011). It is evident that the focus has indeed been on extending quantity rather than improving quality. The NER model considers the number of words in the respoken text (N), the number of edition errors caused by strategies applied by the respeaker (E) and the number of recognition errors that are usually caused by mispronunciations, mishearing or errors with the speech recognition technology (R). These errors can in turn be classified as minor, standard or serious. The threshold for a set of intralingual live subtitles to be considered acceptable is 98%. The need for human intervention is highlighted through the two additional elements of the model: correct editions (CE), which account for editing that has not caused a loss of information, and the final assessment, in which the evaluator can comment on issues such as speed, delay and flow of the subtitles.

---

[1] This section draws heavily upon *Quality assessment in interlingual live subtitling: The NTR model* (Romero-Fresco & Pöchhacker, 2017)

The NTR model (Romero-Fresco & Pöchhacker, 2017) considers the number of words in an interlingually respoken text (N), the translation errors (T) and the recognition errors (R) to calculate the accuracy rate. Thus, the NTR model uses a NER-based formula and accounts for the shift from intralingual to interlingual live subtitling by replacing edition errors (E) with translation errors (T). The latter are in turn subdivided into content (omissions, additions and substitutions) and form (correctness and style) errors. As in the NER model, errors are also classified according to three degrees of severity (in this case minor, major and critical) and the minimum accuracy rate required is 98%.

3.1. Applying the NTR Model

Respoken texts must meet an accuracy rate of 98% to be suitable for broadcast. The user must compare the original audiovisual text with the target text of subtitles to identify each translation and recognition error and effective edition:

- Translation errors must be identified with the error sub-types listed above and then penalised depending on their severity. Three categories of severity highlight the error as either recognisable (*minor*, -0.25), causing confusion or loss of information (*major*, -0.5), or introducing misleading information (*critical*, -1).
- Recognition errors must be identified and penalised depending on their severity as outlined above.
- Differing text that has condensed information or has introduced synonyms is not penalised, rather highlighted as Effective Editions (EE).

The NTR model was applied for the first time in February 2017 to calculate the accuracy rates for this pilot study. An extract of how the model has been applied to the analysis of a respoken text can be seen in Table 3. Examples of translation and recognition errors are highlighted in red and effective editions are highlighted in yellow. Once the criteria above had been applied to all changes between the original audiovisual text and the respoken subtitles, sums of translation and recognition errors were totalled and the NTR formula was applied to calculate the accuracy rate of the respoken text. An example of this can be seen in Table 4.

Table 3.

Extract of NTR Analysis of the Narration Text for Participant 2

Original text (transcribed audio)	Respeaking-based subtitles	Errors
My name is Mary Alice Young. When you read this morning's paper you may come across an article about the unusual day I had last week. Normally there is never anything news worthy about my life, but that all changed last Thursday. Of course, everything seemed quite normal at first, I made breakfast for my family, I performed my chores, I completed my projects, I ran my errands.	Me llamo Mary Alice Young. Cuando lees (leas) (1) el periódico de esta mañana, quizá te encuentres un artículo sobre 10 en (el día) (2) inusual que tuve la semana pasada. Normalmente, no hay nada nuevo en mi vida, era (pero) (3) todo cambió el juez (jueves) (4) pasado. No (todo) (5) era normal al principio, (hice) (6) el desayuno para mi familia, hice mis tareas domésticas, completé mis pequeños proyectos, y e hice mis recados. La verdad, me pasé	1. MinR (0.25): incorrect tense but it does not impact comprehension. 2. MajR (0.5): The target text becomes incoherent and as the error is strange the viewer would not be able to identify the source text. 3. MinR (0.25): This slightly deters from the meaning of the text. 4. MinR (0.25): The source text could still be recognised, given the idea has been mentioned previously.

Table 4.

Extract of NTR Assessment of the Narration Text for Participant 2

Accuracy rate	
MinT: (4 x 0.25 = 1)	MinR: (5 x 0.25 = 1.25)
(cont-omiss) x 1	MajR: (2 x 0.5 = 1)
(cont-sub) x 1	CritR: (1 x 1 = 1)
(form-corr) x 2	
MajT: 0	
CritT: 0	
Total: 1	Total: 3.25

	NTR accuracy rate
N = 243 (209 + 34)	$\dfrac{243 - 1 - 3.25}{243} \times 100 = 98.25\%$ (5/10)
EE: 3	

Assessment
The quality of the subtitles is acceptable. The translation is good (only four translation errors, one regarding content and two regarding style) and perhaps too many recognition errors (10), of which two cause the viewers to lose information and another one introduces misleading information. Still, most errors are minor and therefore do not have significant impact on comprehension.

4. Quantitative Results

An accuracy rate was calculated for each respoken text of the pilot study. Below, a breakdown of errors is provided for each participant and displayed as follows for translation errors: T (MinT – MajT – CritT); and for recognition errors: R (MinR – MajR – CritR). It should be noted that the participant who respoke from Spanish into English is participant 7.

"Video 1: narration" had an average score of 6.2 translation errors and 9.3 recognition errors per participant. The overall average accuracy rate is 97.35% (3/10). The most common errors were content omissions with an average of 9.5 per text, followed by content substitutions (2), form correctness (1.9), form style (0.4) and content addition errors (0.1).

Table 5.

Translation and Recognition Errors per Participant for Video Clip 1

	Translation Errors	Recognition Errors	Accuracy Rate
Participant 1	(1–0–0)	(8–5–0)	97.98% (4/10)
Participant 2	(4–0–0)	(5–2–1)	98.25% (5/10)
Participant 3	(4–0–0)	(8–6–0)	96.93% (2/10)
Participant 4	(3–0–0)	(10–3–0)	97.87% (7/10)
Participant 6	(4–1–1)	(4–4–0)	97.38% (3/10)
Participant 7	(2–4–0)	(3–5–0)	96.42% (1/10)
Participant 8	(9–3–1)	(0–1–0)	97.45% (3/10)
Participant 9	(8–2–2)	(3–1–0)	96.50% (1/10)
Average	(4.4–1.25–0.5)	(5.8–3.4–0.1)	97.35% (3/10)

"Video 2: speech" had an average score of 6.5 translation errors and 4.5 recognition errors. The overall average accuracy rate is 97.38% (3/10). This clip was deemed the easiest clip to respeak. The most frequently made error was content omissions with an average of 3.6 per text, followed by content substitutions (1.1), form style (0.8), form correctness (0.6) and content addition errors (0.1).

Table 6.

Translation and Recognition Errors per Participant for Video Clip 2

	Translation Errors	Recognition Errors	Accuracy Rate
Participant 1	(1–1–0)	(7–5–0)	97.86% (4/10)
Participant 2	(4–0–0)	(3–1–0)	98.75% (6/10)
Participant 3	(4–2–2)	(4–1–0)	97.17% (2/10)
Participant 4	(5–0–2)	(1–3–0)	97.86% (4/10)
Participant 6	(7–2–1)	(0–1–0)	97.58% (3/10)
Participant 7	(3–1–0)	(1–0–1)	96.09% (0/10)
Participant 8	(5–6–0)	(0–1–0)	96.99% (2/10)
Participant 9	(4–1–1)	(2–2–2)	96.73% (1/10)
Average	(4.1–1.6–0.8)	(2.3–1.8–0.4)	97.38% (3/10)

Participant 2 performed exceptionally well compared to the others and had a previous training background in subtitling and respeaking. Participants 1 and 4 performed very well in comparison to the others. Participant 1 had a background in subtitling, interpreting and respeaking and participant 4's background was in subtitling and interpreting. Participants who scored between 95%–97% are not considered to have performed well, as the threshold for live subtitles of good quality is 98%.

Table 7.

Overall Individual Performance with Translation and Recognition Errors

	Translation errors	Recognition errors	Accuracy rate
Participant 1	(2–0–0)	(15–10–0)	97.92% (4/10)
Participant 2	(4–0–0)	(8–3–1)	98.50% (5/10)
Participant 3	(8–2–2)	(12–7–0)	97.05% (2/10)
Participant 4	(8–0–2)	(11–6–0)	97.86% (4/10)
Participant 6	(11–3–2)	(4–5–0)	97.48% (3/10)
Participant 7	(5–5–0)	(4–5–1)	96.26% (0/10)
Participant 8	(14–9–1)	(0–2–0)	97.22% (3/10)
Participant 9	(12–3–3)	(5–3–2)	96.61% (1/10)

The average accuracy rate for the SMART pilot study was 92.78% (0/10), compared to the average accuracy rate for this pilot study being 97.37% (2/10). It would be unfair to compare the two as the videos for this pilot study were approximately 2 minutes long, while those for the SMART study were around 7 minutes. SMART participants were trained for a total of 8 hours, whereas the participants in this study were trained for 2 hours. This may further highlight the impact that speed had on the accuracy rates. In this study, participants would not have experienced much fatigue and had breaks in between respeaking the videos, which could be a contributing factor to higher accuracy rates.

4.1. Translation and Recognition Errors

The most common translation errors were content omissions with an average of 6.6 errors per text, followed by content substitutions with an average of 1.6 errors, form correctness (1.3), form style (0.6) and content additions (0.1). This suggests omissions should become a focal point in interlingual respeaking training. A high number of omissions may indicate that participants either struggled to keep up with the speed of the text or did not know how to translate some parts of the text and resorted to making omissions. A strategy to keep up with the text is to edit or condense it. However, in the case of live subtitling and if editing is not performed correctly, it may cause loss of information for the viewer and in turn lowers the accuracy rate of the text.

An average of 5.8 minor omissions and 0.75 major omissions were made per text, suggesting that participants were omitting pieces of information but they managed to keep the main idea of sentences. As per the NER and NTR models, the loss of a dependent idea unit is the omission of part of a sentence containing information about the "where", "what", "when", "who" and "why" piece of information and is usually scored as a minor translation error. In some instances,

the participant has missed a verb so the *what* piece of information from the sentence is missing. An omission of an independent idea unit is the omission of a whole sentence and is usually scored as a major translation error as it causes substantial loss of information. Most of the time, enough information has been displayed and the viewer can follow what is happening, however, they can still miss information that a hearing audience would have access to.

Results of the SMART study show that content omission and substitution errors were the most common errors, followed by form correctness, form style and finally content additions. Participants also found omissions and substitutions difficult to manage when respeaking. This highlights that managing speed, multitasking and dealing with dense information should be developed within a training programme. The results from the SMART study also point to interlingual respeaking as feasible and ambitious but not unattainable (Davitti, Sandrelli, & Romero-Fresco, 2018).

For this pilot study, participants made on average 6.3 recognition errors in each text, with 8.3 errors in the narration and 4.3 for the speech. For the SMART study, the average number of recognition errors was 5.4 errors per text and more omissions, substitutions and form correctness errors were made than recognition errors. Participants in the SMART study could have made fewer errors as they received more dictation and intralingual respeaking practice. Such practice would have allowed participants to train their voice profiles before beginning interlingual respeaking exercises.

Participants in this study made on average 4.2 minor errors, 1.4 major errors and 0.6 critical errors per text, suggesting they were able to manage the severity of errors. These results suggest more dictation practice is needed to train the user voice profile, dictate at a steady pace and volume and make regular pauses to release words on screen. Participant voices and on-screen actions can be observed in the Screencast clips. Some clips show that poor pace and dictation caused recognition errors. For future studies, more dictation practice is required to allow those who are good live translators not to underperform due to poor dictation or using an untrained voice profile. Participants did not pause enough to release the respoken words on screen. Therefore, they did not have the opportunity to monitor their respoken output and attempt to correct recognition errors live. Again, this points to the importance of building upon and developing new software skills within a training programme.

Participant 1 made the most recognition errors with an average of 12.5 per text and the least amount of translation errors, suggesting they focussed on translation but in turn compromised recognition. In contrast, participant 8 made on average 1 recognition error and 12 translation errors per text, demonstrating they focussed on dictation and compromised translation. There is no doubt that interlingual respeaking is a complex task. Trainers and students must find a happy medium between translation and recognition errors and explore strategies and techniques to overcome both types of error to produce interlingual live subtitles that meet the quality standard.

4.2.   Translation vs. Recognition Errors

Those who achieved 97% or above have an average of 8.5 minor errors, 4 major errors and 0.7 critical errors (penalisation of -4.7). Those who achieved below 97% have an average of 13 minor errors, 8 major errors and 3 critical errors (penalisation of -10.25). This suggests the best respeakers from this experiment were able to keep critical errors to a minimum and managed to have half the number of major errors. For instance, in the case of omissions this means participants managed to omit part of the sentence instead of the full sentence on twice as many occasions. Translation errors can be controlled by the respeaker, whereas recognition errors tend to be uncontrollable when not caused by pronunciation or dictation issues, as they can also be caused by software or hardware malfunction. Therefore, to a certain extent, recognition errors can also be put down to luck.

The comparison between translation errors and recognition errors shows another interesting pattern. The four best respeakers collectively made more recognition errors (63) than translation errors (36), whereas the respeakers who didn't reach 97% collectively made more translation errors (64) than recognition errors (39). The best respeakers of this experiment have very good skills for live translation, as they have fewer and less severe translation errors. Most of the errors these participants have are recognition errors, which means that their scores could have been considerably higher had they received thorough software training. Of the bottom four respeakers, the number of translation errors (64) suggests that live translation was their main weakness, which could perhaps be remedied with extensive language and interpreting training.

5. Qualitative Results

After completing the exercises, participants were asked to rate their level of difficulty on a scale of 1–5 (1 *being easy* and 5 *being difficult*). Participants' self-perception of their performance does not appear to match their actual performance. Those who rated themselves as "satisfactory" scored higher than those who rated themselves as "poor", except for participant 3. This suggests participants were aware of their performance. For example, some reported they felt the software was giving them instant feedback in the form of recognition errors as they respoke. Some participants found this difficult to deal with while trying to perform well. The table below shows the average score of level of difficulty is 4, which shows that interlingual respeaking is perceived as a complex task. It must be noted that students would not have been aware of what score constitutes as poor, good etc. due to a lack of knowledge of the NTR model.

In the post-experiment questionnaire two participants stated the respeaking tasks were linguistically difficult, seven found the speed difficult and one deemed the content comprehension difficult. Other comments included: dealing with long sentences, fear of missing information, their feelings as a respeaker and monitoring their own output.

Table 8.

Participants' Self-Rated Performance Compared with Their Actual Performance

	Level of difficulty	Perception of overall performance	Actual overall performance
Participant 1	5	Satisfactory	97.92% (4/10)
Participant 2	4	Satisfactory	98.50% (5/10)
Participant 3	3	Satisfactory	97.05% (2/10)
Participant 4	4	Satisfactory	97.86% (4/10)
Participant 6	3	Satisfactory	97.48% (3/10)
Participant 7	4	Poor	96.26% (0/10)
Participant 8	4	Poor	97.22% (3/10)
Participant 9	5	Poor	96.61% (1/10)

In the post-experiment questionnaire, participants stated that an interlingual respeaker would likely come across the following challenges: the speed of the source text, remembering to enunciate punctuation, using the software, recognition errors, paying attention to the interpretation, subtitles appearing on screen, linguistic knowledge and multitasking. Participants noted linguistic knowledge and multitasking as the two main challenges for an interlingual respeaker, thus suggesting that an advanced level of the working languages is essential to produce interlingual live subtitles. Multitasking was highly regarded by participants as a skill. This perhaps indicates a requirement for experience of simultaneous interpreting to listen in one language while speaking in another, not forgetting the added element of working with software to correct errors and cue subtitles.

Participants identified various skills that interlingual respeakers would require to improve performance. Overall, these skills were noted in both questionnaires indicating participants correctly predicted the skills deemed necessary to perform well. The following required skills were noted: a strong level of comprehension in the source language, a strong level of expression in the target language, communication, speed, multitasking, listening, software knowledge, memory, segmentation and reformulation. The following qualities were noted as being useful for an interlingual respeaker: ability to work under pressure, ability to keep pace, and focus.

After completing the exercises, six out of eight participants identified that an interpreter would be the best-suited profile for an interlingual respeaker. Two added simultaneous interpreting as the specific mode of interpreting, one noted it should be an interpreter with training in respeaking and another noted an interpreter that can work with the software. Two suggested a translator would be ideal and nobody noted a subtitler as the best-suited professional profile. Some participants noted more than one best-suited profile. Although this experiment yields interesting results on the feasibility of interlingual respeaking, further research is needed to draw meaningful

conclusions. Experience in translation or subtitling along with experience of respeaking are clearly an advantage, but interpreting skills are expected to be the main feature of interlingual respeaking.

6. Training

Given the hybrid form of the task, a training model for interlingual respeaking should be centred around the skills necessary for subtitling, interpreting and intralingual respeaking. A practical proposal for intralingual respeaking was put forward by Arumí Ribas and Romero-Fresco (2008), who note that identifying skills is a fundamental first step for designing a respeaking course. A similar taxonomy of skills may be required to highlight the skills required for interlingual respeaking to inform training. Pöchhacker and Remael (forthcoming) describe interlingual respeaking as a three-step process including pre-process, peri-process and post-process tasks. Dividing skills as per the processes that make up interlingual respeaking would clarify why each skill is necessary.

Interlingual respeaking is about providing a service to heighten access. Therefore, it is important that training is made accessible for trainees. Trainees may be current undergraduate or postgraduate students aiming to learn a new skill, others may be already well-established translators, subtitlers and interpreting professionals. The demographic of participants that took part in this pilot study is likely to be representative of future interlingual respeaking trainees (current postgraduate students and language professionals). An online training programme would cater to the need for accessibility and allow students to work at their own pace. This would not only foster greater flexibility, but also empower students to take control of their own learning, which university students and already established professionals may be more inclined to participate in. From a social constructivist approach, Kiraly (2000) proposes that translator education should be a dynamic, interactive process. A process that is based on learner empowerment and encourages interpersonal collaboration in which teachers serve as guides, consultants and assistants. A training model based on this approach, as explained by Kiraly, would build a sense of responsibility toward their own learning and future profession.

A list of criteria must be drawn up to base the training model upon. Criteria will be guided by the results of both this pilot study, the large-scale study, questionnaire responses and by taking industry training requirements into account. It will also be informed by the NER and NTR models, which outline criteria for quality assessment models. A brief example of criterion is to ensure the model is adaptable and considers language combinations, fast paced change in working environments, industry needs, audience needs and the evolution of interlingual respeaking. Once results of the large-scale study have been examined, other concepts and models for interpreter and translator training must also be considered extensively. Gile's *Effort models of interpreting* (2015) outline a set of models that explain the difficulty of interpreting, which facilitate the development of strategies for better interpreting performance. A set of three main efforts are identified: listening and analysis, speech production and short-term memory. These theoretical concepts can be taught

to allow students to identify difficulties in interpreting and develop strategies to alleviate such difficulties. Due to the similarities between Simultaneous Interpreting and interlingual respeaking, skills such as listening, source text comprehension, speech production and short-term memory are required for both tasks. Therefore, existing research on interpreting pedagogy and the training of intralingual respeakers can both be applied to research and training in interlingual respeaking. A sequential model of translation (Gile, 2015) details the translator or interpreter's progression from the source language to the target language. The model focusses on the comprehension loop and the reformulation loop to verbalise a translation. Gile explains that although translators may have days to find translation solutions, simultaneous interpreters would only have a few seconds. The concepts of this model can be practically applied to interpretation; therefore, such concepts may also be applied to interlingual respeaking given the similarity of the tasks.

Data on translation and recognition errors from this pilot experiment is valuable for deciding on material and resources for further training and experiments. Participants struggled more with content errors than they did with form errors. Content errors are easier to make and refer to the broader content of the text, but form errors refer to the grammar and register of the text, which for many comes naturally. A training programme for interlingual respeaking should focus on the sub-types of translation errors, with an extra focus on managing speed to deal with omissions and managing the content of the text to deal with substitutions.

Audiovisual materials of different genres, topics, speed and content must be selected for a training course. If chosen carefully and set in order of difficulty, they could contribute to students' progress. Processing and reformulating information are necessary in between listening and speaking and are skills that can be built up over time and with each interlingual respeaking exercise. Starting off with videos with a reduced speech rate and long pauses would allow students to grasp the initial skill required to listen and speak at the same time. Then, allowing the speech rate to increase with each respeaking exercise would give students a sense of progress and would allow them to master recognition errors with speed, which are far easier to grasp than translation errors. A course could begin with documentaries, as the slow pace, long pauses and visual images on screen may aid beginners to focus on producing a live translation without feeling rushed. Slow speeches with non-specialised content could also be used, such as those that can be sourced from the EU speech repository for interpreting practice. Then, move on to working with more complex speeches, sports, the news and weather and eventually chat shows with a very high speech rate and multiple overlapping speakers.

This pilot study has identified areas to be targeted within training to master and combat recognition errors. Participants only had a brief intralingual dictation practice of 10 minutes to familiarise themselves with the speech recognition software. However, more extensive practice is required to train the user voice profile and allow students to monitor and correct recognition errors. A unit dedicated to dictation practice would give enough time for students to work on intralingual dictation, interlingual dictation (sight translation) and develop skills in reading in one language

and respeaking in another. This could be considered as an initial step to take before approaching exercises with audiovisual texts. Being able to manage the speech recognition software (including dealing with technical issues, training Dragon with vocabulary, creating macros and error correction) and required hardware (use of the microphone, knowledge of how much the computer can handle at the same time) are also important factors.

6.1.   Applying the NTR Model

To continue some thoughts on training with regards to the NTR model, its primary function is to assess the quality of interlingual live subtitles. However, the model could also be used within a training programme to raise awareness of how live subtitles are assessed. The NTR model could allow students to assess each of their respoken texts, identify errors and categorise their severity. This may provide further understanding of how and why translation and recognition errors are made. It could give students a space to reflect on their performance with different types of audiovisual text and pave the way for students to develop techniques to manage the severity of errors or avoid them altogether. There may also be room for peer review within training to reduce the subjectivity of students analysing their own texts. Peer review may be an effective way of receiving feedback, which could be a turning point in training. Like the NER model, the NTR model allows for an overall qualitative assessment, which could prove insightful for critical peer review exercises between trainers and students alike.

A text that has reached an accuracy rate of 95% could be almost unintelligible but those not familiar with the model could misinterpret accuracy rates. An acceptable accuracy rate for both intralingual and interlingual live subtitles is 98%. To mitigate against any confusion, the NTR model offers a recalculation of the accuracy rate on a 10-point scale. This would serve as another benefit of using the NTR model within a training programme to highlight the actual quality of respoken texts from an early stage in a course and allow students to grasp the concept of accuracy rates quickly.

Subjectivity in scoring has been noted as a minor concern of the NTR model. When assessing translation errors and distinguishing between them, subjectivity could potentially threaten accuracy and consistency. However, previous testing of ten evaluators using the NTR model has already proven to be successful with a low average discrepancy of 0.3 on a scale from 1 to 10 (Romero-Fresco & Pöchhacker, 2017). The average discrepancy of the analysis of this pilot study is 0.38. The author of this article served as the first marker of the texts and a co-creator of the model served as the second. Most discrepancies were due to the content of translation errors, which could be due to the first marker translating out of their mother tongue. If, in an NTR analysis, a minor error has been scored as a major error it will have little impact on the final accuracy rate. For this pilot study such discrepancies minimally increased accuracy rates by an average of 0.12%, which changed the overall score by one point out of 10 only 12% of the time.

This pilot study was conducted in February 2017, so it is believed to be the first experimental study to have applied the NTR model for the analysis of interlingual respoken texts. A total of 18 texts were produced and analysed throughout this pilot study. The model was applied to a small volume of short texts of different types and proved to provide a simple and thorough method of assessing translation and recognition errors. The average number of words for each respoken exercise in this study was 261 words. Texts that achieved 97.50% or above took on average 21 minutes to analyse with the NTR model and texts that achieved 97.49% or less took on average 33 minutes to analyse due to the extra errors that had to be identified, categorised and scored. The simplicity and thoroughness of the NTR model allows for it to be flexible in its application to texts produced by different means i.e. by respeaking and automatic speech recognition; as well as supporting training in interlingual respeaking to analyse the accuracy of interlingual dictation (sight translation) to support the early training of interlingual respeakers.

7. Conclusions

The results presented in this article demonstrate that interlingual respeaking is feasible. Challenges could be overcome by developing task-specific skills through a training programme for interlingual respeakers to build upon skills used in subtitling, interpreting and respeaking. For subtitling, knowledge of SDH, segmentation, reformulation and edition is required. Developing short-term memory, speed and multitasking all highlight the requirement for elements of training to mirror simultaneous interpreting. Specific skills for respeaking would be software related and should include the unlearning of skills, such as speaking in a pleasant voice due to the importance of dictation and enunciation. Live translation skills are essential for interlingual respeaking as is the ability to dictate accurately to the speech recognition software.

Results from this pilot study have shown that an awareness of omissions and recognition errors should be incorporated into a training programme. Omissions have proved to have an impact on overall accuracy rates and cause loss of information for the viewer. Causes of recognition errors highlight dictation as an essential part of respeaking training as many errors were due to participants over-dictating by pronouncing individual syllables of words and causing misspellings of short words. Managing the error severity is an essential aspect of interlingual respeaking. Participants with good live translation skills can control the severity of translation errors by keeping major and especially critical errors to a minimum, even at the expense of increasing minor errors. The severity of recognition errors cannot always be controlled. This emphasises the need for extensive dictation practice and developed software skills to minimise such errors. Participants who are not strong live translators may find it difficult to reach the minimum accuracy threshold of 98%, even if their dictation is good.

The NTR model has proved to be an effective method of assessment for the quality of interlingual live subtitles. It has the potential to be incorporated within a training programme for students to

understand the differences between each sub-type translation error and become aware of what each of their translation choices mean within the overall quality of interlingual live subtitles. Measuring quality during training could prove to be a turning point in training by raising awareness of students' output and the impact it has on the quality of a text that a viewer receives.

More research is required to determine the best-suited professional profile for an interlingual respeaker. The qualitative data suggests an interpreter would be the best-suited. However, the quantitative data shows the highest performing participants did not have previous interpreting training, which may give some hope for subtitling skills to be maintained in training. At this stage, it would be fair to conclude that interlingual respeaking is feasible. This is providing a suitable training programme is produced to train interlingual respeakers to produce quality subtitles and ensure interlingual respeaking becomes a common practice to cater for a DHOH audience and to aid the integration of foreigners into society. Now that this article on the feasibility of interlingual respeaking has been completed, a forthcoming article will focus on the next stage of this project and present the results of a large-scale study. The large-scale study delivered short online training courses to mainly Spanish natives in the UK and Spain. Data, which is in the final stages of analysis, has been gathered for 50 participants from subtitling and interpreting backgrounds. It aims to identify the task-specific skills and best-suited professional profile for an interlingual respeaker. It is hoped that a final article based on the last stage of this project will present a research-informed training model based on the data of the large-scale study. Reporting on the individual stages of research aims to thoroughly document each one, which allows the results to be suitably shared and ensures accurate progression of interlingual respeaking as the most recent mode of AVT.

Acknowledgements

References

Arumí Ribas, M., and Romero-Fresco, P. (2008). A practical proposal for the training of respeakers 1. *The Journal of Specialised Translation*, 10, 106–127. Retrieved from https://www.jostrans.org/

Dawson & Romero-Fresco, (forthcoming). Towards research-informed training in interlingual respeaking: an empirical approach. *The Interpreter and Translator Trainer, 14*. Retrieved from https://www.tandfonline.com/toc/ritt20/current

Davitti, E., Sandrelli, A., & Romero-Fresco, P. (2018). Interlingual respeaking: An experimental study comparing the performance of different subject groups. *Understanding Media Accessibility Quality (UMAQ)*, Barcelona, 4–5 June 2018.

Eugeni, C. (2008). A Sociolinguistic approach to real-time subtitling: Respeaking vs. shadowing and simultaneous interpreting. In J. Cynthia, K. Bidoli, & E. Ochse (Eds.). *English in international deaf communication*. (pp. 357–382). Bern: Peter Lang.

Fresno, N. (2019). Of bad *hombres* and nasty women; the quality of the live closed captioning in the 2016 US final presidential debate. *Perspectives*, *27*(3), 350–366. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/0907676X.2018.1526960?journalCode=rmps20

Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.

Gile, D. (2015). Effort models. In F. Pöchhacker (Ed.). *Routledge encyclopaedia of interpreting studies*. (pp. 135–137). London/New York, NY: Routledge.

Greco, G. M. (2016). On accessibility as a human right, with an application to media accessibility. In A. Matamala & P. Orero (Eds.). *Researching audio description: New approaches*, (pp. 11–33). London: Palgrave Macmillan Limited.

Kiraly, D. (2000). *A social constructivist approach to translator education: empowerment from theory to practice*. Manchester: St Jerome Publishing.

Neves, J. (2004). Language awareness through training in subtitling. In P. Orero (Ed.). *Topics in audiovisual translation* (pp. 127–140). Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.

Pöchhacker, F. & Remael, A. (forthcoming). New efforts? A competence-oriented task analysis of interlingual live subtitling. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18.

Robert, I., Schrijver, I., & Diels, E. (forthcoming). Live subtitlers: Who are they? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18.

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. Retrieved from https://www.tandfonline.com/toc/ritt20/current

Romero-Fresco, P. (2012). Quality in live subtitling: The reception of respoken subtitles in the UK. In A. Remael, P. Orero, & M. Carroll (Eds.). *Audiovisual translation at the crossroads: Media for all 3* (pp. 111–131). Amsterdam/New York, NY: Rodopi.

Romero-Fresco, P. (2015). *The reception of subtitles for the deaf and hard of hearing in Europe*. Bern/Berlin/Bruxelles/Frankfurt am Main/New York/Oxford/Wien: Peter Lang.

Romero-Fresco, P. & Martínez, J. (2015). Accuracy rate in live subtitling: The NER model. In J. Díaz-Cintas & R. Baños (Eds.). *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Hampshire/New York, NY: Palgrave Macmillan.

Romero-Fresco, P. (2016). Accessing communication: The quality of live subtitles in the UK. *Language & Communication*, *49*, 56–69. Retrieved from https://www.sciencedirect.com/science/journal/02715309

Romero-Fresco, P. & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR model. *Linguistics Antverpiensia, New Series: Themes in Translation Studies*, *16*, 149–167. Retrieved from https://lans-tts.uantwerpen.be/index.php/LANS-TTS

Romero-Fresco, P., Melchor-Couto, S., Dawson, H., Moores, Z., & Pedregosa, I. (forthcoming). Respeaking certification: Bringing together training, research and practice. *Linguistica Antverpiensia*, *New Series: Themes in Translation Studies*, *18*.

Szarkowska, A., Krejtz, K., Dutka, Ł., & Pilipczuk, O. (2018). Are interpreters better respeakers? *Interpreter & Translator Trainer, 12*(2), 207–226. Retrieved from https://www.tandfonline.com/toc/ritt20/current