





SMART PROJECT DESCRIPTION

MOTIVATION FOR THE STUDY

The recent explosion of audiovisual content worldwide has made access a high priority to ensure a wide audience can enjoy it. The SMART project addressed these communication needs and accessibility challenges by exploring a new practice, interlingual respeaking, which combines advances in speech recognition technology with human interpreting and subtitling skills to deliver real-time speech-to-text services across languages. As a method, it evolved out of intralingual respeaking, which has become standard practice for broadcasters worldwide in response to legal requirements to produce live subtitles in the same language for people who are deaf or hard-of-hearing. The UK pioneered the development of this media accessibility service. By adding a translation component, interlingual respeaking could revolutionise how subtitles are produced for live foreign-language content, thus making audiovisual content accessible across sensory and language barriers.

Based on its most recent definitions, interlingual respeaking is, in its process, a "form of simultaneous interpreting" (Romero-Fresco & Pöchhacker, 2017), and in its product, "text displayed on a screen with the shortest possible delay" (Romero-Fresco, 2011). This entails two types of shifts, i.e., interlingual (from language A to language B) and intermodal (i.e., from spoken to written). The shift from oral input to written output is enabled by the interaction between humans (the respeaker rendering audio input in the target language) and machines (speech recognition turning the spoken utterances into visual text). Interlingual respeaking extends the challenge of simultaneous interpreting as the respeaker has to convey the source message live to a speech recognition engine in a different target language while obeying the requirements of subtiling, such as speaker identification and additional content labels, adding punctuation orally, controlling prosody to minimise recognition errors, reformulating and monitoring the text thus produced for comprehensibility and readability and editing it, where necessary. Interlingual respeaking thus capitalises on speech recognition technology but combines human input with human real-time translation abilities to optimise efficiency and ensure the quality of the final product (i.e., translated subtitles).

As an emerging practice, interlingual respeaking has the potential to make an ever-increasing amount of multilingual content accessible to foreign and hearing-impaired audiences in various settings, including television, conferences, and other live events. It also offers a way to tackle the growing demand for interlingual live subtitling in the UK and worldwide. As research on interlingual respeaking is still in its early stages, many questions around its viability, the accuracy of the product (i.e., interlingual subtitles) and the necessary competences required by language professionals need further exploration and empirical grounding.

Consequently, empirical data to support the implementation of this method with language professionals was needed as little was known about the required profile of interlingual respeakers, nor how findings in this respect could inform the content of an upskilling course for language professionals. It is these gaps that SMART aimed to fill. SMART extended the pilot carried out in preparation for the main study by working with language professionals from backgrounds relevant to interlingual respeaking (namely interpreting, subtilling, translation and intralingual respeaking) and broadening the analytical scope of the study to account for cognitive and interpretional variables which may play a role in how this technique is learnt and delivered. SMART focused on interlingual respeaking in its human-centred form, i.e., with a human respeaker translating content live and interacting with speech recognition technology to turn the spoken input into written output in a different language.







PROJECT AIMS AND TOPICS COVERED

The SMART project's core was investigating interlingual respeaking as a complex practice through different lenses, both in terms of its process (i.e., how technique-relevant skills are acquired and implemented) and product (i.e., in terms of accuracy of the final translated subtitles). The methodology for the main experiment, which involved fifty-one language professionals with backgrounds in relevant language-related practices, was partially developed and tested during a pilot project that focused on twenty-three postgraduate students with training in interpreting, subtilling and/or intralingual respeaking (Davitti and Sandrelli 2020).

Research Questions and Objectives of the Study

Our research program had three main areas of focus: process, product, and upskilling.

Under process, we aimed to refine the multifactorial models of competence and answer three key questions:

- What human variables contribute to the performance of language professionals?
- What challenges are commonly faced during performance?
- To what extent can performance be sustained?

For the product aspect, we aimed to refine our understanding of what contributes to output accuracy in interlingual respeaking in the performance of language professionals after 25 hours of upskilling. To this end, we addressed three key questions:

- How well do language professionals form different walks of life do after 25h of upskilling?
- Which variables predict (high/low) output accuracy (including professional background, age, cognitive abilities, interpersonal traits as well as error types)?
- How do different individual and content characteristics impact on accuracy?

Lastly, in order to optimise the upskilling of language professionals, we sought to understand:

- What challenges arise during skills acquisition?
- What are the strengths and weaknesses of the upskilling course?

KEY FINDINGS (please see relevant publications for further details)

Selected insights into the process (Research question 1): We investigated empirically (baseline) what cognitive abilities and interpersonal traits underlie interlingual respeaking performance, particularly those that predicted high interlingual respeaking performance. We were also interested in understanding how a 5-week training course on IRSP would affect specific cognitive abilities of the language professionals by comparing their cognitive performance between time point 1 (before the training course) and time point 2 (after the training course). Our primary focus was on executive functions (Miyake et al., 2000), mainly working memory (WM), switching skills, and sustained attention (known effects on simultaneous interpreting as a closely related practice). We used six cognitive tasks (Verbal Fluency Task, Plus-Minus Task, Digit Span Task (simple WM), Reading Span Task (complex WM), N-back Task, and Sustained Attention to Response Task (SART). Firstly, WM was found to predict high interlingual respeaking performance ($\beta = .36$, p = .01). None of the other predictors (i.e., simple WM, N-back, switching skills, and sustained attention) were significant predictors of high performance as p > .05. Additionally, it was confirmed that WM, F(1, 46) = 4.0, p = .05 (from M = .83, SE = .02 to M = .88, SE = .02) and switching skills, F(1, 49) = 6.42, p = .02 (from M = 22.90 s, SE = 2.95 s to M = 14.55 s, SE = 1.85 s) were enhanced after the upskilling course,







indicating that these skills can be trained and are malleable. However, sustained attention did not improve as p > .05, indicating that alternative forms of attention should be investigated in any further studies (e.g., divided attention). Eight scales were used to measure different interpersonal traits: trait anxiety, resilience, impulsivity, cognitive flexibility, innovativeness in IT, personality, work motivation, and mindfulness. We found that conscientiousness (TIPI), $\beta = -.32$, p = .02 and integrated regulation (WEIMS) $\beta = -.28$, p = .04) negatively predicted accuracy. The other areas of interpersonal traits were not significant. These findings can be linked to this practice's real-time and cognitively demanding nature.

Selected insights into the product (Research question 2): We used the NTR model (Romero-Fresco & Pöchhacker 2017) to measure the accuracy of 153 performances (three per participant) under different scenarios (speed, planned/unplanned delivery, and multiple speakers). An intelligibility scale (based on Tiselius, 2009) was applied to determine high and low performers of interlingual respeaking, which was validated in the results obtained. Across all participants and testing scenarios, the average accuracy was 95.37%. (for reference, the intralingual respeaking accuracy benchmark is 98%; there is no established benchmark for interlingual respeaking yet). The results per language directionality were: Romance into English: 96.16% and English into Romance: 94.89%. In relation to scenarios, the overall results were the following: for speed 94.76%; for planned/unplanned delivery 95.83%; for multiple speakers 95.51%. This shows how speed was felt as the most challenging scenarios – although it is important to bear in mind that all conditions can be found in any given source speech (hence the calculation of average accuracy across all testing scenarios to start with). In terms of errors, across all scenarios, omissions ($\beta = -1.12$, p < .001) were the strongest negative predictor of accuracy, followed by substitutions ($\beta = -.17$, p < .001) and recognitions ($\beta = -.34$, p < .001) .001). Effective editions, i.e., changes in the target text, which did not lead to a loss of information, were positive predictors of accuracy across all scenarios ($\beta = .31$, p = .03). High performers (n = 27) scored significantly higher (M = 96.3%, SE = .2%) than low performers (n = 24, M = 94.4%, SE =.2%) across all scenarios, p < .001. They made significantly fewer omission and correctness errors and used effective editions significantly more than low performers.

Selected insights into upskilling (Research question 3): Traditionally, interlingual respeaking has been taught as a separate set of skills after intralingual respeaking, typically in full modules that cover various aspects like intralingual respeaking, simultaneous interpreting, pre-recorded subtitling as distinct entities. However, this practice involves a unique combination of skills that may not align precisely with those required for other practice, despite some overlaps. In other words, interlingual respeaking is a distinct skill set with its own specific demands, even though it shares some similarities with other language-related practices. In SMART, we wanted to test a more flexible and modular approach to upskilling for language professionals from different backgrounds who may already possess some relevant skills and may need to adjust/unlearn or acquire others. We, therefore, broke down the procedural skills needed into subskills that were taught progressively and could be practised separately (e.g., simultaneous listening and speaking/translating, software-adapted delivery, punctuation, chunking strategies, dealing with speed, software optimization, and error correction), before bringing them all together. Another innovative aspect of our course was the integration of intralingual practice as a stepping stone toward interlingual respeaking. This involved adding the translation component gradually, with alternating practice between intralingual and interlingual respeaking tasks. Additionally, we designed the training to confront participants with challenging scenarios. They were tested on long video clips (approximately 15 minutes) that simulated real-world situations likely to pose difficulties during performance. This approach was implemented in upskilling course at the core of our experiment. Findings from our research questions on the process and product of interlingual respeaking have enabled us to confirm the effectiveness of various aspects of our design. For example, we gained insights into the attainable levels of accuracy following just 25 hours of training (overall, as well as per language directionality and scenario), which are quite impressive for such a brief training duration. We also identified common error patterns, which errors had statistically more impact on accuracy (such as omissions, substitutions and recognitions) and some key challenges arising during the course, for example handling time lag, live error detection and correction, audiovisual monitoring and software-adapted delivery. From performance observation and







qualitative comments, we could also obtain information on coping strategies naturally developed by participants during the course, as well as indications on what to focus on more (for instance, intelligent summarization to reduce omissions and content loss; specific tasks such as cognitive shadowing and related reformulation, especially when associated with software-adapted delivery; high-speed speeches, using speaker identification macros to cope with rapidly alternating speakers and developing multitasking skills). Additionally, we introduced an innovative approach to grouping participants that reflects their composite professional background, thus aligning with the reality of the language industry. Findings confirmed that participants from all profile groups (spoken-to-spoken, spoken-to-written, and mixed background) can acquire interlingual respeaking skills but may need to focus on different aspects. For instance, participants belonging to the spoken-to-spoken group (i.e., mostly used to working with spoken language) found it particularly challenging to get used to software-adapted delivery instead of natural intonation and monitor the output instead of looking at the speaker and dealing with corrections. Participants belonging to the written group (i.e., mostly used to working with written language) seemed to find it more challenging than other groups to get used to real-time translation and keep up with the pace. Finally, participants from the mixed background group (i.e., used to working with both spoken and written language) identified as key challenges handling delay, chunking, and inserting punctuation. These insights highlighted areas where participants needed more emphasis and practice and provided very useful indications for further modularised and tailored upskilling. 82% of participants found that the course met their initial expectations, providing many comments to support their answers. Despite finding the course challenging (86%), they commented on the progress they made throughout the course; even those who felt initially frustrated felt that their performance improved throughout the course (90%). In summary, our study showed that the upskilling approach developed has several advantages; supporting learning, in that by teaching subskills separately first and then gradually bringing them together, participants could build competence in individual components before attempting the full interlingual respeaking task. Teaching subskills progressively also helped some participants manage their cognitive load and perceive the learning process as manageable despite its complexity. This approach also enabled our participants to pinpoint exactly what skills they need to adjust, unlearn or acquire from scratch, thus highlighting the importance of flexibility, modularity, tailoring and personalisation on the basis of varying skills levels and backgrounds. Based on these findings, the upskilling course is currently being refined as part of the SMART-UP project funded by the ESRC Impact Acceleration account (PI https://www.surrey.ac.uk/research-projects/smart-shaping-multilingual-access-Davitti. 2023-24. through-respeaking-technology-upskilling).