



Interlingual live subtitling: the crossroads between translation, interpreting and accessibility

Luis Alonso-Bacigalupe¹ · Pablo Romero-Fresco¹

Accepted: 25 July 2023

© The Author(s) 2023

Abstract

Whereas prototypical conceptions of translation and interpreting as two separate and distinct entities are still being upheld in different contexts, the professional reality shows a less clear-cut picture. Given the relentless advancement of technological innovation and the important role played by (media) accessibility in society, translation, interpreting and accessibility are in a state of flux, with increasingly blurring boundaries between them. Interpreting is not only concerned with spoken and signed modalities, but also with written target texts, which links it to audio–visual translation. In turn, audio–visual translation is no longer only about dubbing and subtitling films from one language to another. It now covers the provision of intra- and interlingual translation for live programmes and live events for users who have no access to the original version (be it for linguistic, sensory-related or other reasons). Speech-to-text interpreting (STTI), also referred to as live subtitling, is a communication-enabling service that allows the production of written access to live events or programmes for people with and without hearing loss. It thus brings together translation, interpreting and accessibility. The aim of this paper is to introduce STTI for readers who may not be familiar with it and to delve into interlingual live subtitling (ILS), and more specifically to explore the research carried out to date about this new technique, its professional development and the way in which it is being trained.

Keywords Accessibility · Speech-to-text interpreting (STTI) · Interlingual live subtitling (ILS) · Intralingual and interlingual respeaking · Automatic speech recognition (ASR) · Machine translation (MT)

1 Introduction

As contemporary societies become increasingly globalised, technologised, diverse and inclusive (at least in theory), forms of communication are also changing. Translation and interpreting, a discipline concerned with communication across cultures and languages, is no stranger to this state of flux. Until now, the boundaries between the two areas were kept clear by training institutions such as universities, most of which distinguish between modules and subjects about translation and those devoted to interpreting. The same goes for standardisation bodies such as the committees in charge of the ISO standards for interpreting, which define the latter as the “rendering of spoken or signed information

from a source language to a target language in oral or signed form” (ISO 20212018:8). There is no room for any interaction between interpreting and the written word in the target language.

However, reality is showing a much more hybrid scenario, as shown in Fig. 1, produced by Pöchhacker and Romero-Fresco for the EU-funded project Interlingual Live Subtitling for Access (ILSA). The contents of the figure will be discussed below, but suffice it to say now that the production of subtitles for live TV programmes/events, known as speech-to-text interpreting (STTI) or live subtitling, bears witness to a healthy and vibrant relationship between interpreting and the written word, operating at the crossroads between (audio–visual) translation and media accessibility.

The aim of this paper is to introduce STTI or live subtitling as a (relatively) new form of communication that can help bring down established barriers (Sect. 2) and to provide an analysis of how interlingual live subtitling is currently being approached in terms of research (Sect. 3), professional practice (Sect. 4) and training (Sect. 5).

✉ Luis Alonso-Bacigalupe
lalonso@uvigo.es

Pablo Romero-Fresco
promero@uvigo.es

¹ Department of Translation and Linguistics, Universidade de Vigo, Galicia, Spain

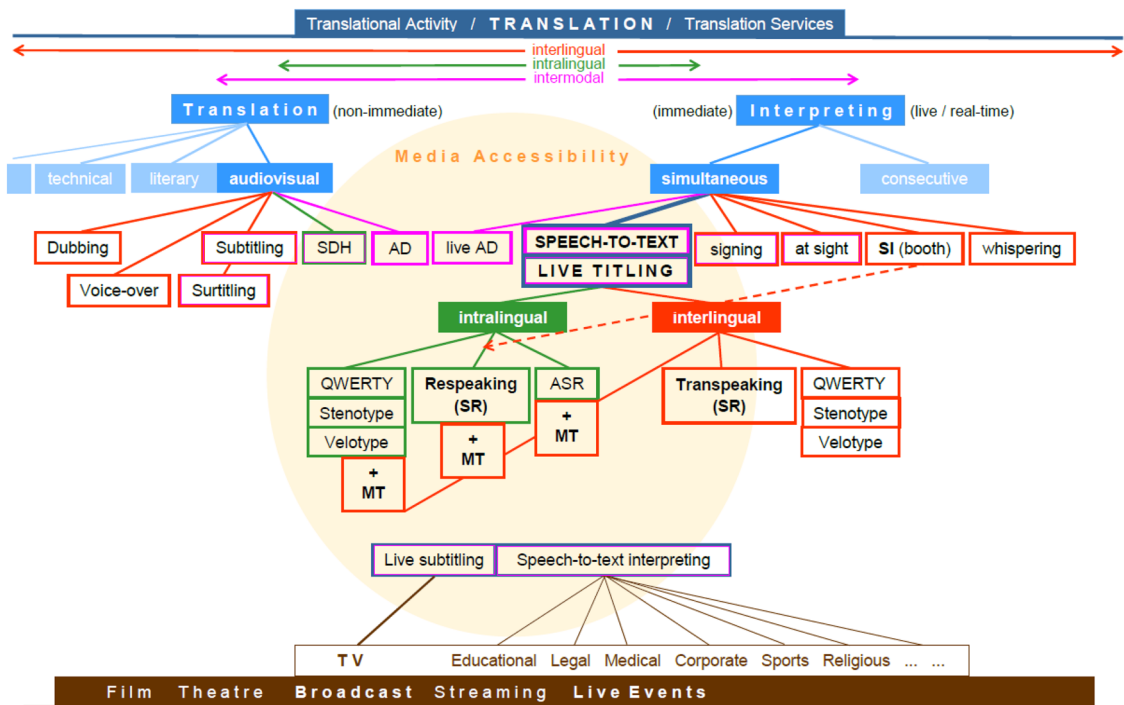


Fig. 1 Position of speech-to-text interpreting (live (sub)titling) in the translation, interpreting and media accessibility map

2 What are STTI and ILS?

STTI, known in different countries and contexts as live subtitling or live/real-time captioning, involves the production of a written version of a spoken message while it is being delivered. It is a form of diamesic translation [1], that is, a change in language mode, in this case from speech to writing. STTI has a relatively long tradition as an intralingual communication service for deaf and hard-of-hearing people [2], but it is now also used interlingually in a wide range of settings (e.g. TV, conferences, workplace, political, educational), event types (e.g. breaking news, business meetings, parliamentary debates, classroom interaction, museum tours, etc.) and formats (e.g. one speaker, multiparty interaction, etc.) [3]. In STTI, an interpreter listens to the original soundtrack of a programme or event and produces a written transcription (either edited or verbatim, but in the same language) or a written translation (into another language) using a keyboard, a stenotyping machine or speech recognition software. The latter option requires respeaking, that is, an interpreter (respeaker) who dictates or translates what is being said, adding punctuation marks, to the speech recognition software, which displays the output as text on screen [4].

As noted above, when produced intralingually, STTI does not consist of a mere process of repetition or transcription. STT interpreters are often forced to edit and

rephrase the source text (for instance because they cannot keep up with the speaker's speech rate) and are normally expected to add information for viewers with hearing loss, such as identification of different speakers and descriptions of sounds. Until recently, and with the exception of some European countries where live events are subtitled intralingually through keyboards, most intralingual live subtitling for either TV programmes or live events has been produced through stenotyping (especially in countries such as the USA and Canada) or respeaking, which has taken over as the most common method [5]. Over the past few years, some companies and TV stations are also beginning to test and use fully automatic subtitles, which use speech recognition technology to transcribe the speaker's voice directly, without intervention from an interpreter [6].

As for interlingual STTI or interlingual live subtitling (ILS), the topic of this paper, it can also be done by combining different techniques and professionals with various degrees of human-machine interaction, for example, in a hypothetical scenario of English into Spanish translation:

- Interlingual respeaking: a single STT interpreter listens to the source text in English and dictates it into Spanish to a speech recognition software;
- Simultaneous interpreting and intralingual respeaking: a simultaneous interpreter translates the English audio into Spanish audio and an intralingual respeaker turns the Spanish audio into Spanish subtitles;

- Simultaneous interpreting and automatic speech recognition (ASR): a simultaneous interpreter translates the English audio into Spanish audio and an ASR engine turns the Spanish audio into Spanish subtitles;
- Intralingual respeaking and machine translation (MT): an intralingual respeaker turns the English audio into English subtitles and a MT engine turns the English subtitles into Spanish subtitles;
- A fifth method or workflow is the combination of ASR (to transcribe English audio into English subtitles) and MT (to translate English subtitles into Spanish subtitles). This is a fully automatic method that does not involve any direct human intervention.

3 Research on STTI and ILS

Research studies comparing the efficiency of different ILS workflows are only now beginning to take off, but there are relevant precedents whose origins can be traced back to the end of the past century.

A pioneering experiment was conducted in 1988 by Kurz and Katschinka [7] involving the simultaneous interpreting of English audio into German audio and the subsequent production of German subtitles by a professional subtitler. Den Boer [8] and de Korte [9] presented the results of a similar experience in the Netherlands in the 1990s with the participation of two teams of two subtitlers: one would do the simultaneous interpreting of the verbal message from English into Dutch, while the other subtitler would type the subtitles using a special type of keyboard known as Velocity. In both cases, the results showed that one major hurdle for the audience was an excessive delay between the production of the source text and the moment in which the subtitles were shown on the screen [10].

3.1 Recent studies

One of the most recent analyses of the quality of different ILS methods was conducted by Eugeni in 2020 [11] during the course of an international conference. Using this natural environment enabled Eugeni to provide insight into a number of real-life but non-comparable situations from an experimental standpoint, with source texts in different languages and translation in different directions, but with very interesting implications for the purpose of this paper. Eugeni tested 5 different workflows: (1) simultaneous interpreting and stenotyping, (2) interlingual velotyping, (3) intralingual velotyping and MT, (4) intralingual respeaking and MT and (5) ASR plus MT plus live editing. His results, included in Table 1, show that the most efficient workflow is 1, whereas the least efficient one is 3, followed by 5. This latter finding

Table 1 Results of Eugeni's (2020) research on ILS: speed (WPM), accuracy (IRA) and delay

	ILS 1	ILS 2	ILS 3	ILS 4	ILS 5
WPM	110 (20.4%)	114 (17.8%)	141 (3.3%)	118 (14%)	132 (4.5%)
IRA	97.3%	95.8%	71.2%	92.1%	86.9%
Delay	7.3"	4.8"	4.3"	6.8"	8.1

may be seen as somewhat unexpected, as this is the only workflow that requires three agents (or resources).

In 2021, Daniela Eichmeyer-Hell [12] compared the quality (in terms of accuracy and user preference) of the intralingual subtitles in German provided by respeakers and typists in a natural environment as well. Although Eichmeyer's study did not include the analysis of interlingual subtitles, it may still be considered pertinent, as one of the workflows she analysed (intralingual respeaking) is very relevant for the purpose of this article. Eichmeyer looked at the quality of the subtitles produced by professional subtitlers, either using a conventional keyboard or through respeaking. The results showed that the respoken subtitles were 12% more accurate—as measured with the IRA model [12]—than those produced by a conventional keyboard. The participants pointed at the delay between the source text and the subtitles as one of the main aspects to be improved, a reminder of the fact that effective communication is not just a matter of linguistic accuracy, but rather the result of a combination of factors that make up the overall users' experience.

Also in 2021, Pablo Romero-Fresco and Luis Alonso-Bacigalupe [13] put to the test five workflows for the production of ILS. One of the distinctive characteristics of this study is that, as in the case of Eichmeyer's, the participants were experienced professionals, including two conference interpreters with more than 20 years of experience in the field, two professional intralingual respeakers and two professional interlingual respeakers. Romero-Fresco and Alonso-Bacigalupe tested the efficiency (in terms of accuracy, delay and cost) of the following five ILS workflows: (1) interlingual respeaking; (2) simultaneous interpreting and intralingual respeaking; (3) simultaneous interpreting and ASR; (4) intralingual respeaking and MT; and (5) ASR and MT. The experiment was done entirely online and involved one single language combination (EN–ES). Accuracy was measured using the NTR Model [14]. The results showed that three of the workflows (1, 2 and 4) were beyond the acceptability threshold of 98% according to the NTR model. However, the analysis of delay and cost yielded a much more nuanced scenario that may limit the potential usefulness and acceptability of some of the workflows. As far as delay is concerned, workflow 5 (the worst in terms of accuracy) ranked first, workflows 1, 3 and 4 provided

intermediate results and workflow 2 (the best in terms of accuracy) ranked last, which is not surprising, as it is the only one which involved the participation of two human resources. This study also looked at the potential cost of the service in terms of the amount of resources required for each workflow. The workflows assessed ranged from those that are fully human (1 and 2) to those that are semiautomatic (3 and 4) and finally one that is fully automatic (5). The analysis shows, as in the case of Eugeni [11], an inversely proportional relationship between automatization of the workflow and cost. The more automatic the workflow, the more affordable it is. Conversely, the higher the number of humans involved, the higher the cost of the service.

Hayley Dawson [15] replicated this experiment using the same workflows and language pair, but resorting to a different language direction (ES–EN), different participants (all professionals, but in this case native English speakers) and different source texts. Unlike in Romero-Fresco's and Alonso-Bacigalupe's study, where three workflows reached the 98% accuracy threshold set by the NTR model, none of the workflows tested by Dawson achieved 98%. Yet, despite the difference in accuracy rates, Dawson's results are very much in line with those obtained by Romero-Fresco and Alonso-Bacigalupe. Workflows 3 and 5 ranked in the last two positions in the accuracy analysis, although in different order, and workflows 1, 2 and 4 were the best in terms of accuracy with very similar results. A few trends may be emerging here. Firstly, the two fully human workflows (1 and 2) ranked amongst the best in both experiments, whereas the two workflows using ASR ranked the lowest in both cases. Secondly, intralingual respeaking rendered very good results, either in combination with simultaneous interpreting or with MT. Finally, interlingual respeaking seems to fare well in both experiments when considering both accuracy and delay.

More recently, Alice Pagano [16] has carried out a double study for her doctoral thesis. In her first experiment, she analysed the accuracy of three workflows previously tested by Dawson and Romero-Fresco and Alonso-Bacigalupe (1, 2 and 3) with the English–Italian language pair. None of the workflows reached the minimum 98% accuracy threshold, with results that were even lower than Dawson's (97.2% for workflow 1, 96.9% for workflow 2 and 95.1% for workflow 3). Yet, the consistency between both studies seems clear, with the three workflows occupying the same positions (interlingual respeaking ranked first, followed by simultaneous interpreting + intralingual respeaking and simultaneous interpreting + ASR). In her second experiment, Pagano assessed the five workflows tested in the previous two studies, in this case with the language pair Spanish–Italian. As found by Romero-Fresco and Alonso-Bacigalupe, workflow 2 obtained the highest accuracy (the only one reaching 98% for Pagano) and workflows 1, 3 and 4 took intermediate

positions with an accuracy rate around 97.1% in all three workflows. In keeping with the trend observed so far, the poorest results were obtained by the fully automatic workflow 5 (ASR + MT), with an accuracy rate of 95.4%.

3.2 Emerging trends

Table 2 presents a comparative view of the results obtained in the experiments discussed in the previous section. Columns 1 and 2 (from left to right) identify the different workflows (except in Eugeni's research, where they have been placed right below the accuracy results in column 3). Columns 3, 4, 5, 6 and 7 include the results (accuracy rate and ranking) obtained by Eugeni [11], Dawson [15], Romero-Fresco and Alonso-Bacigalupe [13] and Pagano's [16] experiments 1 and 2, respectively.

Figures 2 and 3 provide a snapshot of the general trends resulting from the above-discussed findings. Regardless of the specific figures, there seems to be a considerable level of consistency between the results obtained in the different analyses.

Before discussing the first tentative trends resulting from the comparative analysis of these studies, it is important to note that this is not a systematised battery of experiments designed and carried out in parallel by one single group of researchers. Instead, it is a compilation of the results obtained by a number of researchers who share the same objectives and ambitions, but who have worked separately (although not in isolation), with different language pairs and directions, and testing different workflows in different situations, from real-life environments to experimental settings, and from online to face-to-face scenarios. Unsurprisingly, then, the results are not totally consistent in terms of specific accuracy rates, but some interesting trends are beginning to emerge:

- Workflow 2 (simultaneous interpreting and intralingual respeaking) seems to be the most efficient method, ranking first in three of the analyses and second in the other two. Workflow 5 can be found on the other side of the spectrum, ranking 4th or 5th in the four studies in which it was tested;
- Workflow 1 (interlingual respeaking), the only one where one human alone is in charge of the whole ILS process, fared well in the experiments, except in Pagano's study, where it ranked 4th. Eugeni did not test interlingual respeaking, but interlingual velotyping, which achieved very good results. More research is needed to draw clear conclusions, but it does seem that this type of ILS carried out by a single human is neither a pipe dream nor an unsurmountable task for interpreters, but rather a thriving trend and a promising new opportunity with sufficient

Table 2 Accuracy rate and ranking of the ILS workflows tested

ID	Workflows	Eugeni	Dawson	Romero and Alonso	Pagano 1	Pagano 2
1	Speech-to-text interpreting	95.8% ILS2	2	1	97.2%	1
2	SI + intralingual respeaking	97.3% ILS1	1	2	96.9%	2
3	SI + ASR	95.4%	5	97.6%	3	97.2%
4	Intralingual respeaking + MT	92.1% ILS4	3	98.8%	2-1	97.2%
5	ASR + MT	71.2% ILS3	5	97.2%	5	95.4%
Lang. pair and direct Assess. model		86.9% ILS5	4	EN-ES	EN-IT	ES-IT
		Varied	ES-EN	NTR	NTR	NTR
		IRA	NTR			

potential for widespread use as long as good training is provided;

- Workflow 4 (intralingual respeaking and MT), which, as will be discussed in the next section, remains virtually untested in the professional arena, has yielded good results in all studies. Conversely, workflow 3 (simultaneous interpreting and ASR), which is being favoured by language and access service providers, has fared fairly poorly.

As has been discussed, these trends seem all the more promising if we take into account the different conditions under which the various studies were carried out (language pairs/directions, real-life vs experimental settings, online vs face-to-face scenarios, etc.). In this sense, there are two additional factors that may have significant implications for future research on ILS: the profile of the participants and the type of training provided.

The highest accuracy results, including some above the 98% quality threshold set by the NTR model, were obtained in the studies by Dawson [15] and Romero-Fresco and Alonso-Bacigalupe [13], both of which resorted to professionals. Pagano's participants, who were postgraduate students in conference interpreting, did not reach this threshold. Still, their results were fairly good, which may be explained by the length of the training she provided: one 70-h workshop for each of the experiments she conducted. Also relevant is the experience obtained through the SMART project, in which 35 professional subtitlers and interpreters with different language combinations were trained as interlingual respeakers. Although some top-performing participants reached accuracy rates of 96–98% with the NTR model, the average accuracy rate was 95.37%, well below the minimum 98% target. Here, the low average accuracy may be explained by the fact that, despite being professionals, the participants only received unmonitored online training, which did not seem to be as efficient as the training provided in the other studies. Although more research is needed to draw firm conclusions, it would seem that, given the challenging nature of ILS, in order to obtain optimum accuracy rate results in empirical studies in this area, it may be necessary to use professionals and to provide them with as thorough and supervised a training programme as possible.

Finally, as mentioned above, quality and efficiency in ILS are not only determined by accuracy. Other factors, such as delay and cost, must also be factored in. An analysis of the studies that looked at delay (Eugeni, Pagano, Dawson, Romero-Fresco and Alonso-Bacigalupe) and cost (Romero-Fresco and Alonso-Bacigalupe) shows an interesting pattern: the more automatic the method (especially workflow 5), the lower the delay, the cost and the accuracy. Conversely, the more human the method (especially workflows 1 and 2), the higher the delay, the cost and the accuracy. In between

Fig. 2 Accuracy rate of ILS workflows as measured with the NTR and IRA models

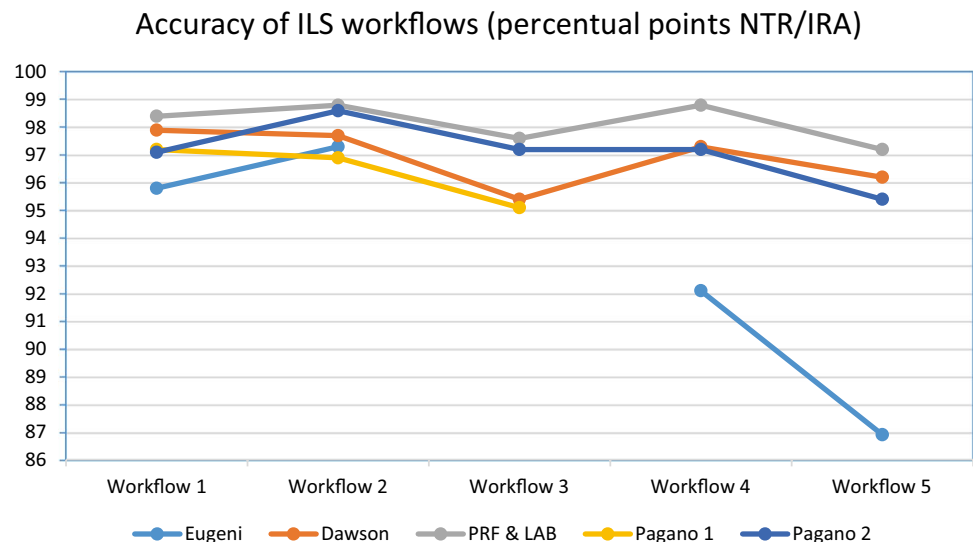
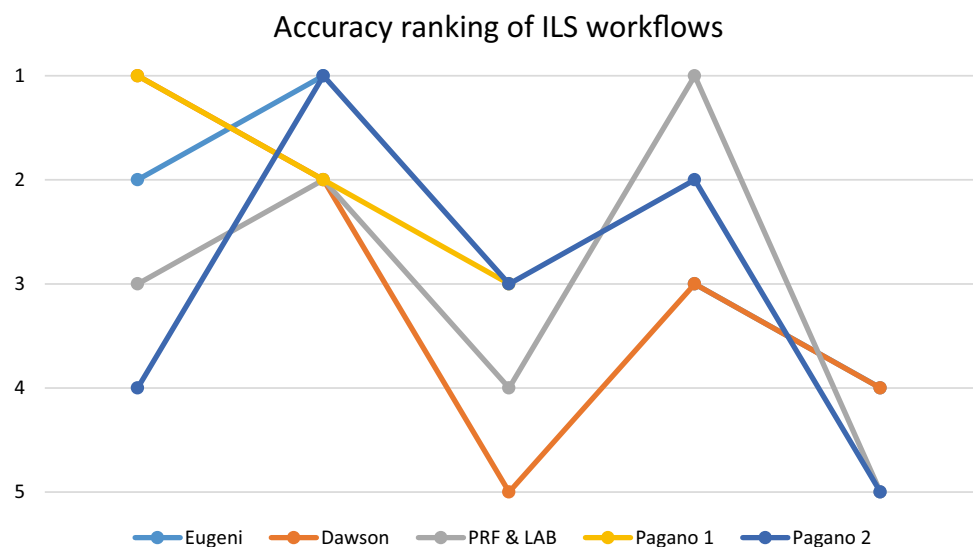


Fig. 3 Accuracy ranking of ILS workflows



these two poles are workflows 3 and 4, which may not be as accurate as the completely human methods but which may constitute an interesting happy medium in terms of overall quality and efficiency. The following section provides a brief account of how the industry is reacting to these different workflows when it comes to the production of ILS.

4 Professional ILS

All five workflows tested in the studies analysed in the previous section are being used by the industry. Some are, however, more common than others. As explained by professional STT interpreter Nancy Guevara [17, 18], the most recurrent ones seem to be workflows 1 and 2: interlingual respeaking and the combination of simultaneous interpreting

and intralingual respeaking. Figure 4 shows Guevara's set-up for the provision of ILS through workflow 1¹:

¹ Guevara's [16] set-up for the provision of ILS through interlingual respeaking (workflow 1) consists of:

“(1) second laptop (2) my co-STTI, Jacobo (3) ultrawide screen (4) run sheet / prep notes (5) open internet tab for quick searches (6) captioning platform: input field into which I dictate; my co-STTI then edits any errors I miss (7) speech recognition software (8) live output to check what the audience is receiving (9) webcam (10) captioning platform control buttons: handover, standby, edit access (11) chat window for communicating with co-STTI and tech support (12) notes: common macros I use in captions, e.g. [UNKNOWN TERM], [INAUDIBLE] (13) video of event streaming live in background, with volume low (14) Zoom call to connect with interpreters, and (15) chat window to communicate with them (16) main laptop, all apps open (17) lots of water (respeaking really dries your throat!) (18) on/off and volume control for headset microphone (19) ergonomic mouse: precision trackball helps to quickly edit errors live (20) notes: event-specific macros (21) printed prep document with relevant terms, such as proper nouns and speaker names (22) ergonomic key-

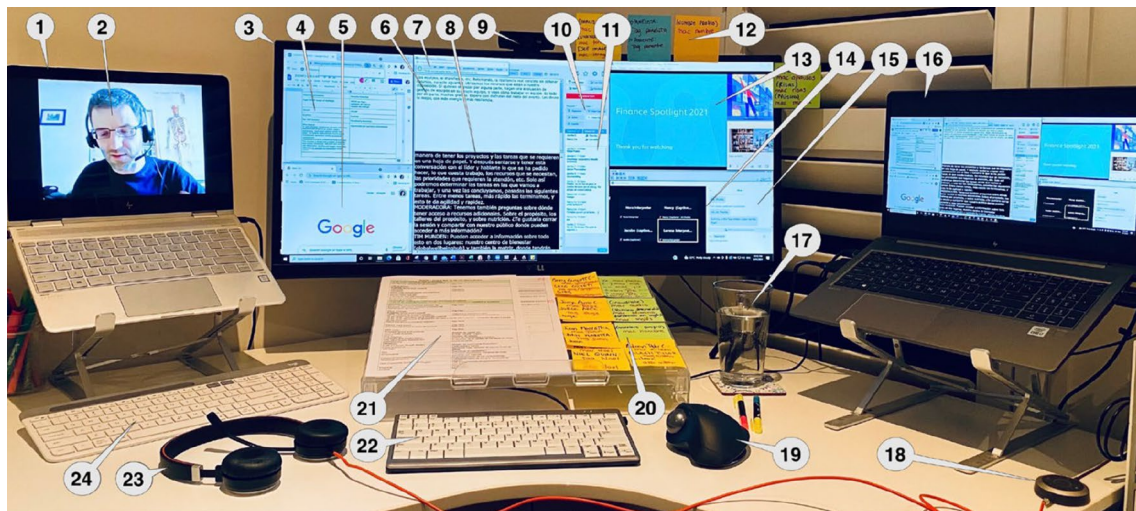


Fig. 4 Nancy Guevara's set-up for the provision of interlingual live subtitles through interlingual respaking (workflow 1)

In a recent article published by *In Touch*, the magazine of the Australian Institute of Interpreters and Translators, Guevara reflects on what it is like to work at the intersection of two different industries: media accessibility and language services. She explains that most of the work currently comes from media access companies specialised in intralingual live captioning/subtitling. The challenge here lies in educating the companies regarding the requirements of simultaneous interpreting, the need for preparation material and the working conditions required for optimum provision of interlingual live subtitles, including swapping with a co-interpreter every 15 min. Language service providers, in turn, need to learn about subtitles and the technology required for their production and provision. Guevara [18] believes that STT interpreters can contribute to bridging the gap between these two industries and “not only help expand accessibility further through multilingual captions, but also maximise the work opportunities created by the demand for this service”.

Workflows 1 and 2 are, however, not the only ones being used in the professional ILS market. Leading companies within the sector are now offering three-tier access provision, which ranges from very high quality and standard cost (workflows 1 and 2) to good quality and reduced cost (workflows 3 and 4) and finally acceptable quality and very affordable cost (workflow 5). Very much in line with a contemporary market-oriented perspective, it is for the customer (conference organisers, broadcasters, etc.) to decide the type of access provision they require and they can afford to have.

An interesting example is that of the EU Parliament. As is well known, each Member of the Parliament has the right to read and write parliamentary documents, follow debates and speak in their own official language. The European Parliament, which pledges to be accessible and transparent for all citizens of the EU, has an in-house translation service to produce the different language versions of its written documents and communicate with EU citizens in all the official languages. It also has interpreting services for multilingual meetings organised by the official bodies of the institution. However, at the moment, Members of the European Parliament cannot access debates on screen in their own language and these debates are not accessible for deaf and hard-of-hearing people.

To tackle this problem, the Official Journal of the European Union published on 6 August 2019 an invitation to tender to acquire a licence for a tool “that is able to automatically transcribe and translate parliamentary multilingual debates in real time” [18], that is, a tool that can provide ILS using workflow 5. The tender rules out human subtitling (that is workflows 1 and 2) because it requires “a high degree of multilingualism” and it is “a highly resource-intensive task” (ibid: 4). It adds that automatic ASR is already being used to facilitate the production of verbatim transcripts of plenary sessions. The new live speech-to-text and machine translation tool goes a step further in order to provide “better, cost-efficient services for cross-lingual communication for its Members and European citizens” (ibid: 4). The tool will start working in 10 core languages (English, German, French, Italian, Polish, Spanish, Greek, Romanian, Dutch and Portuguese) and will then be rolled out to all 24 official languages of the European Parliament.

The implementation of the tool is subject to a positive evaluation of the quality of its output, which will

Footnote 1 (continued)

board (23) headset with microphone (24) back-up keyboard with dual connection to both laptops.”.

be monitored by a team of researchers that includes the authors of this article. This quality assessment is all the more important in light of the above-mentioned arguable results obtained by workflow 5 in the experimental studies conducted so far. This does not mean, of course, that this workflow cannot aspire to obtain optimum results someday in the future, but, at the moment, and considering that it is going to be applied in 24 languages, it seems unrealistic to anticipate a scenario where the results can be fully satisfactory in the short run. Based on the results shown in this article, it would have been interesting to consider workflows 1 and/or 2, which could guarantee high-quality output while also helping to integrate the activities of the translation and interpretation units of the EU Parliament. If workflow 2 (simultaneous interpreting and intralingual respeaking) was to be adopted, simultaneous interpreters could provide an audio translation for the hearing audience and intralingual respeakers working alongside them would produce a written translation for users who have no access to audio or who prefer subtitles. After the speech, these intralingual respeakers could amend any errors in the subtitles and produce perfectly synchronised subtitles for the video of the event. This would integrate translation, interpreting and access services at the EU Parliament, bringing them in line with the hybrid scenario pictured in Fig. 1 and with the professional reality described by Nancy Guevara [17, 18].

5 Implications for training

Perhaps because live subtitling has been so far mainly performed intralingually (as a form of media accessibility and thus not involving language transfer), this technique has so far received much more attention amongst professionals and scholars in subtitling than in interpreting. This also applies to training in live subtitling, which is normally delivered at graduate and postgraduate courses in subtitling or in-house by subtitling companies [4].

Even though intralingual live subtitling, especially through respeaking, was introduced in Europe as a profession in 2001, the provision of formal training at higher education (HE) level did not start until 2007. During this six-year period, and given the lack of research, codes of practice or even basic guidelines, companies had no option but to train their own professionals, many of whom were already working as pre-recorded subtitlers (see Fig. 5, from [5]).

Whereas courses on subtitling for pre-recorded films and TV programmes have proliferated in Europe over the past decade, live subtitling and respeaking training at university level is still scarce. Most of it is exclusively intralingual and can be found in self-contained modules or as part of larger modules within MAs in AVT, where the prerequisite is a translation or language-related BA [5]. Thus, respeaking

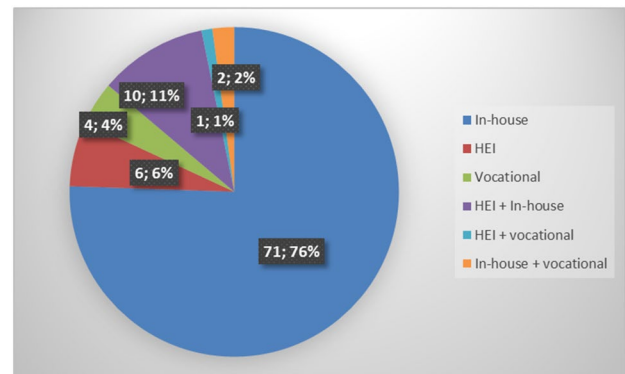


Fig. 5 Type of training for intralingual live subtitling

training at HEIs ranges from introductory sessions on intralingual respeaking as part of postgraduate courses on AVT (such as at the University of Leeds, in the UK, Universitat Autònoma de Barcelona, in Spain, and the University of Parma, in Italy) to the more thorough training offered by the University of Antwerp (a six-month face-to-face course in Dutch), and the University of Roehampton (a three-month face-to-face module in English, Spanish, French, Italian and German), with Universidade de Vigo offering an exclusive online tutored six-month programme on the interlingual modality that will be described below.

Now that the market demands ILS, for which there is very little training available at the moment, it remains to be seen whether this will be taken up by subtitling or interpreting programmes. This is directly related to the question of what profile (subtitlers or interpreters) is best suited to produce ILS.

In her doctoral thesis, Dawson [19] set out to identify the task-specific skills required for interlingual respeaking, which turned out to be multitasking, live translation, dictation, command of source and target languages and comprehension. She also conducted the first comprehensive comparative analysis of the performance of interpreters and subtitlers doing intra- and interlingual respeaking. In general, interpreters obtained better results than subtitlers in both intra- and interlingual respeaking, but there were differences within and across groups. Her results show that, especially when it comes to interlingual respeaking, being an interpreter does not guarantee good respeaking performance, just as being a subtitler does not need to be an obstacle to become a good interlingual respeaker. The good performing interpreters were strong live translators and were able to keep up with fast speeds and to deal with the multitasking aspect of respeaking. Much like the good performing interpreters, good performing subtitlers had clear dictation, good live translation skills and also seemed to keep up with the text. It appears that although interpreters may be better equipped initially to deal with the complexity of interlingual

respeaking, students from other backgrounds may also have the necessary task-specific skills to perform well. This highlights the fact that the most important aspect for trainees may be the development of the task-specific skills that they lack from previous experience rather than having a particular professional profile.

Partly informed by the preliminary results of the research conducted by Dawson [19] for her doctoral thesis and with the aim of developing and testing the first training course on ILS, the University of Vigo, as leader, and the Universities of Vienna (Austria), Antwerp (Belgium) and Warsaw (Poland), as partners, set up, in 2017, the EU-funded three-year ILSA (Interlingual Live Subtitling for Access) project. The ILSA course is a full-fledged and freely available self-training online programme for anyone interested in the field, although it is particularly suited to fine-tuning the skills of subtitlers and simultaneous interpreters and to adjust and adapt them to the principles and specificities of STTI. The course is made up of foundational modules (simultaneous interpreting, pre-recorded subtitling and media and live events accessibility), core modules (intralingual respelling and interlingual respelling) and applied modules (TV and live events and education). This is, however, a self-learning course, which may make it difficult for trainees to become proficient, given the challenging nature of STTI. In view of this and drawing on the contents of the ISLA course, the University of Vigo offers a 200-h fully tutored online postgraduate course in STTI made up of three basic modules: simultaneous interpreting (60 h), intralingual respelling (60 h) and interlingual respelling (80 h). The partner universities in the ILSA project (Antwerp, Vienna and Warsaw) have set up similar courses, which, with very few additions in Italy and Germany, makes for a very limited training provision in STTI that can hardly meet the current demand for this service worldwide.

In our view, this is likely to remain the same until STTI comes out of the strict confines of audio–visual translation/media accessibility and is adopted by the interpreting community as a (relatively) new modality of simultaneous interpreting. In an attempt precisely to place STTI on the interpreting map and to recognise the work of the professionals currently performing this challenging task, a group of practitioners and scholars approached the committee in charge of the official ISO/FDIS 22259 Standard on Simultaneous Interpreting Delivery Platforms to enquire about the inclusion of STTI in the document. Initially, the committee refused to mention STTI in the standards on the grounds that respelling, whether intra- or interlingual, is not interpreting and that there are no good quality examples of STTI out there, two arguments that are disproved by the literature included in this chapter. This negative response was not surprising, especially taking into account that in ISO standards, interpreting is defined as the “rendering of spoken or signed

information from a source language to a target language in oral or signed form” (ISO 20212018:8). In other words, a prototypical (and perhaps obsolete) conception of interpreting that turns its back to the professional reality described by Guevara above, where, beyond well-established definitions, interpreting may also be understood as the rendering of spoken information from a source language to a target language in the written form.

A second reason for the rejection of STTI as a form of interpreting may be that it is perceived as a threat for working interpreters. Yet, as shown in several recent studies [19], interpreters are ideally positioned to become respellers, given that they already possess many of the skills required to undertake the task. Training in STTI can give them another string to their bow so that they can provide different forms of interpreting depending on the context and user requirements. Closing the door to this reality is likely to undermine the development of STTI and to encourage the industry to adopt fully automatic methods that are still not ready to provide high-quality translations. The new accessibility initiative by the EU Parliament, described in the previous section, is a case in point.

Fortunately, recent developments, including the increasing demand for STTI worldwide, have caused the committee in charge of the official ISO/FDIS 22259 Standard on Simultaneous Interpreting Delivery Platforms to set up a working group for the inclusion of STTI in the standard. This may pave the way to the recognition of STTI within the interpreting community and its inclusion in interpreting training programmes. As highlighted by [10] in their detailed analysis, the skills required in STTI make it an even more challenging task than simultaneous interpreting, which would perhaps warrant its inclusion in postgraduate rather than undergraduate programmes. However, intralingual respelling, involving no language transfer and thus considerably less demanding than interlingual respelling, may be a good fit for graduate training. Furthermore, interpreting students would benefit from being exposed to examples of STTI from the beginning of their undergraduate training, so that they understand how translation, interpreting and media accessibility co-exist in a scenario that is no longer compartmentalised, but rather fluid and hybrid.

6 Final thoughts

As the only type of interpreting that is accessible to both hearing people and people with hearing loss, STTI is a particularly timely and useful service in a society that is currently striving to be more diverse and inclusive. Yet, the limited inroads STTI has managed to make so far into the interpreting community, especially when it comes to overall recognition and training provision, show how difficult it is to

do away with prototypical conceptions that keep firm boundaries between translation, interpreting and accessibility.

However, as is often the case, it is the professional market that sets the demand (and the pace), which may explain why researchers and trainers are now beginning to jump on the STTI bandwagon. This article has provided the first comparative analysis of the research carried out in this area so far with regard to the efficiency of the different workflows currently available for the provision of ILS. Despite the differences across the studies, some interesting patterns have emerged. The methods that have the most human input (workflows 1 and 2) are the ones that yield the highest accuracy, but also the highest delay and cost. Conversely, the more automatic the method (especially workflow 5), the lower the accuracy, the delay and the cost. As a happy medium in between these two poles stand workflows 4, which has obtained very good results, and 3, which has so far fared poorly.

Given the significant differences in cost (and delay) between the ILS workflows, it is not surprising that more and more companies and institutions are testing semi- or fully automatic methods. Researchers and scholars have here an important role to play, as current evidence shows that, for instance, workflow 5, although tempting from a financial point of view, falls very short of providing the quality standard required.

Drawing on the results obtained in [1], we have pointed to a scenario of “horses for courses”, where the choice of a particular workflow over the others may be based on aspects such as the level of accuracy required, the type of delay that can be considered acceptable and the financial means of the customer. However, the comparative analysis included in this article shows that some systematic trends are beginning to emerge, which leads to the recommendation of workflows 1, 2 and perhaps 4. All three provide new and exciting opportunities for interpreters and subtitlers who wish to branch out and expand their professional portfolio. That said, speech recognition and machine translation are improving at a steady pace and it is not unrealistic to envisage that workflows 3 and, perhaps to a lesser extent, 5 may yield better results in the future.

A brief reflection may be in order here. The development of these workflows, especially workflow 5, may be seen as a threat to interpreters and subtitlers. Yet, we should note that our allegiance as researchers lies with the quality of the service provided for the users. Monitoring, assessing and, whenever possible, improving this quality are the ultimate goal; however, the service is produced. Be that as it may, the current scenario shows that the human methods yield the best quality and even if semi- or fully automatic methods manage to raise to the same standard, human input will likely be needed to assist and supervise technology, not to mention to ensure that the provision of subtitles is tailored

to the needs of different types of users [20]. The future looks vibrant and promising for those wishing to develop a career in this area. Here is hoping that higher education institutions embrace the opportunity to expose trainees to the different ILS workflows and to train them so that they can deliver this service as a new form of multilingual communication that brings together the affordances of translation, interpreting and accessibility.

Acknowledgements This research has been conducted within the framework and with the support of the Spanish government-funded project “The Quality of Live Subtitling (QuaLiSub): A regional, national and international study” (PID2020-117738RB-I00) and the ALMA—Network for research excellence in media accessibility.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Gottlieb, H.: Semiotics and translation. In: Malmkjær, K. (ed.) *The Routledge Handbook of Translation Studies and Linguistics*, pp. 45–63. Routledge, London (2017)
2. Stinson, M.S.: Speech-to-text interpreting. In: Pöchhacker, F. (ed.) *Routledge Encyclopedia of Interpreting Studies*, pp. 399–400. Routledge, London (2015)
3. Davitti, E., Sandrelli, A.: Embracing the complexity: a pilot study on interlingual respeaking. *J Audiovis Transl* **3**(2), 103–139 (2020)
4. Romero-Fresco, P., Eugeni, C.: Live subtitling through respeaking. In: Bogucki, Ł., Deckert, M. (eds.) *Handbook of Audiovisual Translation and Media Accessibility*, pp. 269–297. Palgrave, London (2020)
5. Robert, I.S., Schrijver, I., Diels, E.: Live subtitlers: Who are they? A survey study. *Linguist. Antverpiensia New Ser. Themes Transl. Stud.* **18**, 101–129 (2019)
6. Pérez Cernuda, C.: Subtitulado automático bilingüe: una solución no tan sencilla. *Panorama Audiovisual.Com* (2022). <https://www.panoramaaudiovisual.com/2022/10/25/subtitulado-automatico-bilingue-la-idea-es-sencilla-la-solucion-no-tanto/>
7. Kurz, I., Katschinka, L.: Live subtitling: a first experiment on Austrian TV. In: Nekeman, P. (ed.) *Translation, our future:*

- Proceedings of the XIth World Congress of FIT. Euroterm, Maastricht, pp 479–483 (1988).
8. Den Boer, C.M.: Live interlingual subtitling. In: Gambier Y, Gottlieb H (eds.), (Multi)media translation: Concepts, practices, and research, pp 167–172. John Benjamins, Amsterdam (2001). <https://doi.org/10.1075/btl.34.20boe>
 9. De Korte, T.: Live inter-lingual subtitling in the Netherlands: Historical background and current practice. InTRAlinea Special issue: Respeaking (2006). http://www.intralinea.org/specials/article/Live_interlingual_subtitling_in_the_Netherlands
 10. Pöchhacker, F., Remael, A.: New efforts: A competence-oriented task analysis of interlingual live subtitling. *Linguist. Antverpiensia New Ser. Themes Transl. Stud.* **18**, 130–143 (2019)
 11. Eugeni, C.: Human-computer interaction in diamesic translation. Multilingual live subtitling. In: Dejica, D., Eugeni, C., Dejica-Cărțiș, A. (eds.) *Translation studies and information technology - new pathways for researchers, teachers and professionals*. Translation Studies Series. Editura Politehnica, Politehnica University Timișoara, pp. 19–31 (2020).
 12. Eichmeyer-Hell, D.: Speech recognition (Respeaking) vs. the conventional method (Keyboard): A quality-oriented comparison of speech-to-text interpreting techniques and addressee preferences. In: Jekat, S.J., Puhl, S., Carrer, L., Lintner, A. (eds.) *Proceedings of the 3rd Swiss Conference on Barrier-free Communication (BfC 2020)* (2021).
 13. Romero-Fresco, P., Alonso-Bacigalupe, L.: An empirical analysis on the efficiency of five interlingual live subtitling workflows. *XLinguae*, special issue edited by Bogucki Ł & Deckert M. <https://doi.org/10.18355/XL.2022.15.02.01> (2022).
 14. Romero-Fresco, P., Pöchhacker, F.: “Quality assessment in interlingual live subtitling: the NTR model. *Linguist. Antverpiensia New Ser. Themes Transl. Stud.* **16**, 149–167 (2017)
 15. Dawson, H.: Interlingual live subtitling: a research-informed training model for interlingual respeakers to improve access for a wide audience. Doctoral Thesis. Department of Media, Culture and Language. University of Roehampton (2020).
 16. Pagano, A.: Testing quality in interlingual respeaking and other methods of interlingual live subtitling. Doctoral Thesis. Department of Modern Languages and Cultures. Università di Genova (2022).
 17. Guevara, N.: Speech-to-text interpreting Part 1: new skills for a T&I portfolio. *Ausit-In touch Mag. Austral. Inst. Interpreters Transl.* **29**(2), 11–12 (2021)
 18. Guevara, N.: Speech-to-text interpreting Part 2: Speech-to-text interpreting, part 2: some practical insights. *Ausit-In touch Mag. Austral. Inst. Interpreters Transl.* **29**(3), 16–17 (2021)
 19. Dawson, H.: Exploring the quality of different live subtitling methods: a Spanish to English follow up case study. Paper presented on 17 September 2021 at the 7th IATIS conference held at Pompeu Fabra University (Barcelona, Spain) (2021).
 20. Romero-Fresco, P., Fresno, N.: The accuracy of automatic and human live captions in English. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* **22** (2023).
 21. DGT (Directorate-General for Translation): Live speech to text and machine translation tool for 24 languages - Innovation partnership – Specifications (2019). <https://etendering.ted.europa.eu/cft/cft-display.html?cftId=5249>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.