

## EDITING MACHINE-GENERATED SUBTITLE TEMPLATES: A SITUATED SUBTITLER TRAINING EXPERIENCE

Alejandro Bolaños García-Escribano<sup>1</sup>

<sup>1</sup>Universitat Jaume I & University College London

**Abstract:** Automation technologies have altered media localisation workflows as much as practitioners' workstations and habits. Subtitling systems and streaming services now often integrate built-in automatic speech recognition (ASR) engines, sometimes even combined with machine translation engines, to produce subtitles from audio tracks. The rise of post-editors in the audiovisual translation (AVT) sector, specifically subtitling, has been a reality for some time, thus triggering the need for up-to-date training methods and academic curricula. This article examines the uses and applications of editing practices for machine-generated timed transcriptions in subtitler training environments. A situated learning experience was designed for an international team of eight AVT trainees and three educators to edit raw machine-generated subtitles (both inter- and intra-lingually) for educational videos. The publication of an accessible video book by a publishing house was the ultimate objective of this project, undertaken by an international team of English- and Spanish-speaking postgraduate students and graduates. The feedback collated after this experience through an online questionnaire proved paramount to understanding the use of subtitle post-editing for ASR-produced templates in AVT education. Interestingly, most respondents believed that subtitle post-editing training, be it intralingual or interlingual, should be further embedded in translation curricula while also identifying bottlenecks that AVT educators may find useful when developing activities of this nature. **Keywords:** Audiovisual translation; Subtitling; Automation technologies; Automatic speech recognition; Post-editing



## 1. Introduction

Our society experiences an ever-growing daily consumption of audiovisual content (Nikolić & Bywood, 2021), especially thanks to the democratisation of streaming platforms and the internetisation of information (Lobato, 2018). New technologies, especially those that can generate content with little human support, such as automatic speech recognition (ASR) and machine translation (MT), are gaining momentum in the media localisation industries (Burchardt *et al.*, 2016; Díaz-Cintas & Massidda, 2019). Professional systems can now produce machine-generated, auto-timed subtitle templates from audio input employing ASR to further automatise the subtitling process, in which dedicated automation technologies are becoming a staple tool (Mehta *et al.*, 2020).

In the age of artificial intelligence and the fourth industrial revolution (Schwab, 2017), the rise of professionals who can edit texts before and after the use of automation tools in media localisation, specifically in subtitling, is a reality (Georgakopoulou & Bywood, 2014; Bywood, Georgakopoulou & Etchegoyhen, 2017). Subtitlers face increasing volumes of editing work, but scholarly attention has been scarce on this front (Athanasidi, 2017; Koponen *et al.*, 2020). Following the *Machine Translation Manifesto* produced by Audiovisual Translators Europe (2021), training would-be subtitlers in pre- and post-editing is necessary to meet industry demand, which calls for up-to-date training methods that expose students to those technologies and professional editing (Bolaños-García-Escribano & Declercq, 2023).

This article examines a didactic experience in which ASR-generated content was used in a subtitler training setting to localise educational videos for publication. Despite the growing body of literature on translation automation and the teaching of post-editing, most research efforts have been devoted to MT engines (Torrejón & Rico, 2012; Bowker, 2015; Rico, 2017). ASR technologies have featured prominently in fields such as respeaking and live subtitling (Romero-Fresco, 2011) and interpreting (Defrancq & Fantinuoli,

2021). Yet, little research has been carried out on the inclusion of ASR – combined with editing – in the subtitling classroom. Most audiovisual translation (AVT) scholars merely mention the existence of ASR technologies in the industry (Bolaños-García-Escribano, Díaz-Cintas & Massidda, 2021) or its benefits for professional dubbing (Mejías Climent & Lozano, 2021). Their growing importance makes their examination increasingly relevant for subtitle training purposes. This article discusses the results of a pedagogical project in which an international team of eight translation trainees edited raw automatically generated captions and subtitles for an accessible video book entailing nine video presentations (240 minutes). The experience took the form of a situated translation project in which participants worked on subtitle templates for educational videos with the help of experienced translator trainers and staff from an international publishing house.

### **Machine-generated subtitle templates and editing**

Audiovisual texts, made of sounds and images, encompass four different types of signs to produce meaning, i.e., audio-verbal, audio-nonverbal, visual-verbal and visual-nonverbal (Delabastita, 1989), thereby building complex semantic composites (Sokoli, 2005; Zabalbeascoa, 2001). AVT practices constitute different language transfer modes whose linguistic output can be written for subtitling (e.g., interlingual subtitling, captioning) or spoken again for revoicing (e.g., voiceover, dubbing). There has been increasing academic interest in the complex semiotic texture of audiovisual texts (Gambier & Gottlieb, 2001; Pérez-González, 2014, 2018; Bogucki & Deckert, 2020), with many studies discussing translational techniques and reception.

In this article, an emphasis will be put on subtitling, which has many sub-types, depending on whether they are either interlingual or intralingual, open or closed, among other characteristics. Following the development of digital formats, subtitling has been an ally of

globalisation and the preferred mode of AVT on the internet (Díaz-Cintas, 2012, 2023). Automation tools have been progressively integrated into subtitling workflows in the past few years (Burchardt *et al.*, 2016) alongside the use of pre-timed templates, which prompt subtitlers to focus on the translation rather than the spotting of clips (Nikolić, 2015; Georgakopoulou, 2019). Such developments have arguably been particularly useful in turning subtitling into an inexpensive AVT practice for interlingual translation, which is increasingly being used on social media, video games and streaming platforms, as well as in corporate communication and tutorials. According to the latest market surveys (Valuates Reports, 2022), the volume of interlingual subtitles needed for international dissemination should continue increasing; moreover, subtitling comes third among the services most commonly offered by top translation service providers (Hickey, 2023).

Technology has been a prominent ally of all types of translation activity, facilitating the transfer of source-language text into the target version while enhancing productivity and cost-efficiency with the help of computer-aided translation tools (Bowker, 2002; Quah, 2006). In the AVT industry, however, the use of automation tools has been considerably timider, with only a few tools starting to offer MT or ASR tools a couple of decades ago (Georgakopoulou, 2018). AVT-specific systems are used to tackle the technical challenges of revoicing and subtitling practices, namely spotting, visualisation of the sound waveform, shot changes, characters per line and display rates, among others in the case of subtitling, and notations, pauses and character tags in the case of revoicing. However, other than spell checkers, few automation tools have been integrated into mainstream subtitling software (Athanasiadi, 2015, 2017), with most advancements being developed for more agile cloud-based platforms (Bolaños-García-Escribano & Díaz-Cintas, 2020).

Subtitlers' productivity is determined by the efficiency of the technologies used and the time required for the target files to be deemed fit for purpose. In this sense, using automation tools followed by post-editing (and sometimes preceded by pre-editing)

only makes sense if the degree of editing is low and translators can escalate their output. Much research has been devoted to understanding how machine-generated translation output, generally of a lesser quality than human translations (Läubli & Orrego-Carmona, 2017), can be improved and how it affects translators' work in terms of cognition-related aspects such as effort (Moorkens *et al.* 2015). To reach acceptable quality standards, machine-translated raw output has to undergo a human revision of the output, aka *post-editing* (BSI, 2015). Post-editing differs from revision in many aspects, such as the nature of the errors encountered in the translated text and the expected quality level (Mossop, 2020). Often individually elaborated for each institution (Allen, 2003), post-editing guidelines lack homogeneity in the industry, but there seems to be consensus, as explained by Hu & Cadwell (2016), that there are two main types of post-editing depending on how much the output ought to resemble human-generated content; the terms *full* (aka “publishable quality”) and *light* (aka “good enough” or “fit for purpose”) can be found in commonly used post-editing guidelines such as the one published by TAUS (2016).

In AVT scholarship, and as maintained by Matamala (2017), MT and post-editing have received substantial attention, with a number of EU-funded projects yielding a sizable amount of research outputs. The latter include the likes of MUSA (Multilingual Subtitling of Multimedia Content, 2002–2004), eTITLE (European multilingual transcription and subtitling services for digital media content, 2004–2005), SUMAT (Subtitling by Machine Translation, 2011–2014), EU-BRIDGE (2012–2015), CompAsS (Computer-Assisted Subtitling, 2018–2019), MuST-Cinema (Multilingual Speech-to-Subtitles, 2020–), and MediaVerse (A Universe of Media Assets and Co-creation Opportunities at Your Fingertips, 2020–2023), among others. Of particular interest were EMMA (European Multiple MOOC Aggregator, 2014–2016) and TraMOOC (Translation for Massive Open Online Courses, 2015–2018), which focused on the application of MT technologies to localise educational videos from English into eleven languages (Kordoni *et al.*, 2016).

There seem to have been fewer efforts devoted to ASR in interlingual (media) translation, with only a few studies on the post-editing of ASR-generated content, such as Matamala *et al.* (2015), Tardel (2020), and Vitikainen & Koponen (2021). Large-scale projects on the use of ASR technologies in AVT include TransLectures (Transcription and Translation of Video Lectures, 2011–2014), EU-BRIDGE (Bridges Across the Language Divide, 2012–2015), HBB4All (Hybrid Broadcast Broadband for All, 2013–2016), and MeMAD (Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy, 2018–2021). Another scholarly attempt at integrating ASR into AVT was a small-scale project entitled ALST (*Accesibilidad Lingüística y Sensorial: Tecnologías para la audiodescripción y las voces superpuestas*, 2013–2015), which focused on audio description and voiceover and did not show promising results for ASR output “[...] probably due to the testing conditions” (Matamala, 2015, p. 81). Yet, I very much agree with Georgakopoulou (2019, p. 526), who claimed that “[...] ASR certainly has the potential to revolutionize the AVT industry further through improvements and innovations in its use.”

Today, ASR and MT engines seem to be making far-reaching inroads into subtitling, with advanced speech-to-text and transcription tools that offer high accuracy rates, such as AWS (<https://aws.amazon.com/transcribe>), HappyScribe (<https://www.happyscribe.com>), Omniscien (<https://omniscien.com>), Rev (<https://www.rev.com>), and Speechmatics (<https://www.speechmatics.com>). Some companies are developing their own systems – or finetuning their proprietary tools – to further automatise in-house translation labour using ASR combined with MT systems (Mehta *et al.* 2020). In contrast, professional subtitling systems are starting to offer ASR and MT functionalities thanks to API integrations; for instance, MateSub (<https://matesub.com>) and SyncWords (<https://www.syncwords.com>) can generate auto-timed templates in various languages employing ASR and MT, whereas commercial cloud subtitling software programs such as OONA Tools (<https://oona>).

oonatools.tv) now integrate ASR for automatic template creation as well as MT engines – such as AppTek (<https://www.apptek.com>), XL8 (<https://www.xl8.ai>), and Amazon Translate (<https://aws.amazon.com/translate>) – for automatic template translation followed by post-editing.

This article focuses on ASR as a means to produce machine-generated, auto-timed intralingual transcriptions in the form of editable subtitle templates. In this article, the terms *editing* and *post-editing* are indistinctly used to refer to the revision of ASR-generated transcriptions, whereas *pre-editing* was not considered because the ASR technologies used in this study can produce timed transcriptions without any human input. To assess the potential usability of ASR-generated subtitles, however, one has to determine how they meet their purpose, that is, whether the target-language templates can attain good-quality standards both technically (i.e., spotting) and linguistically (i.e., transcription). Despite high accuracy rates reported by many ASR developers, light and full editing will almost always be required for ASR-generated content to be deemed fully acceptable and of professional standard. According to Bolaños-García-Escribano & Declercq (2023, p. 576), there are “[...] six different types of editing that traditionally take place in the AVT industry – i.e., pre-editing, post-editing, revision, proofreading, reviewing QC [quality control] and post-QC viewing.” They also report on the additional editing-related complexity posed by the subtitling of media programmes, and that *truncation* (i.e., reduction of information) can occur at any point when producing subtitles. This type of editing regards the partial or total condensation of information to override spatio-temporal restrictions that are inner to subtitling. Indeed, subtitle post-editors necessarily truncate subtitles when post-editing auto-timed templates to abide by character-per-line and display-rate limitations, among other factors.

## Research methodology and materials

To better understand the potential of using ASR-generated subtitle templates alongside post-editing in the classroom, a situated learning experience was designed. In situated learning, pedagogical practices take place in “[...] authentic, real-world professional settings” (González-Davies & Enríquez-Raído, 2016, p. 1), which can vary depending on each community of practice. This study involved a team of would-be subtitling professionals who had previously received subtitling-specific training during their undergraduate and postgraduate studies. It followed an international e-conference on the role played by media accessibility in translation education and language-learning environments, which took place in 2020. Professional respeakers provided live captions for accessibility purposes, but these could not be exported in subtitle format. To overcome this hurdle, a project was developed between the host institution of the e-conference and the publishing house to undertake the localisation of the video chapters and make them fully accessible to the target readers. The video chapters, preceded by an introductory lecture, were finally published in 2021 by UCOPress in an ISBN-borne video book edited by Bolaños-García-Escribano, Veroz-González & Ogea-Pozo (2021). It was offered in an accessible interactive format (FlipHTML5) and included timed intralingual captions and interlingual subtitles.

Designed to be a collaborative project between UCL’s Centre for Translation Studies and the University of Córdoba, Spain, postgraduate students and recent graduates were allowed to apply for a place in a funded project whose main aim was the localisation of a video book employing intralingual captioning (English) and interlingual subtitling (Spanish to English). The situated nature of this learning experience lies in the fact that the work was carried out following a professional commission (i.e., the intended publication of an accessible video book) of authentic materials (i.e., educational videos) from a real-world client (i.e., the publisher). Those postgraduate translation students who applied



for a position were interviewed and selected based on merit and availability throughout the project (May to September 2020), which was partially funded by UCL with a ChangeMakers grant (<https://www.ucl.ac.uk/changemakers/ucl-changemakers>). The Spanish university's publishing house was ultimately responsible for processing and publicising the project's outcome (i.e., the video book). Those selected to participate were offered monetary retributions for their work. A written agreement outlining the participants' responsibilities and expected outcomes was established between the project members and the participating students. As part of this collaboration, the staff partners involved in the project provided *ad hoc* training on subtitling – including ASR and subtitle templates – and editing, and produced the project's guidelines and teaching materials. They also helped assist and supervise the students and mediate with the client (i.e., the publisher) as and when necessary.

The original clips contained over 240 minutes of video footage (60 minutes in Spanish and 180 minutes, in English). The video book was published in English with intralingual captions. In addition, intralingual and interlingual subtitles were produced for the 60-minute Spanish-language video. A total of eight student collaborators – four Spanish natives and four English natives<sup>1</sup> – were assigned different tasks: three English-speaking project members were responsible for the post-editing of 180 minutes of English-language video content, whereas two Spanish-speaking project members edited the 60-minute Spanish-language video (i.e., intralingual captioning), and two English-speaking participants were

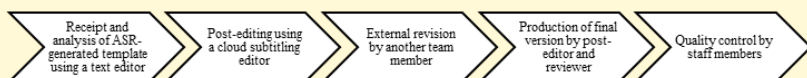
---

<sup>1</sup> The low number of participants is an obvious limitation of this study in terms of how the research results can be extrapolated beyond this small-scale study. However, I would like to argue that situated learning experiences often include a low number of students due to the complexity of academia-industry partnerships. In this particular case, because of the length of the ASR editing tasks, priority was given to providing students with individualised feedback as well as maximising the limited funding available for participants to receive fair monetary retributions for their work.

responsible for the Spanish into English subtitles (i.e., interlingual subtitling). A dedicated shared communication and file-sharing hub was created in a Microsoft Teams channel, in which several built-in applications were used for participants to distribute the work and communicate with each other effectively. The staff members managed the whole workflow remotely, and the different activities were arranged and monitored using a Gantt chart as well as the file-sharing folders within the channel. Participants could report on the status of their ongoing work so that staff members could monitor progress and supervise the workflow by inspecting the deliverables and providing feedback in writing as and when necessary.

The workflow (see Figure 1) was devised and explained to the participants and the rest of the project's stakeholders. Participants were provided with key deadlines and the tools necessary to undertake the work (e.g., subtitling editors, file-sharing tools, communication channels) between May and September 2020. The video books were submitted to the publisher in late 2020 and ultimately published in early 2021.

**Figure 1:** Rough workflow for the post-editing of ASR-generated templates

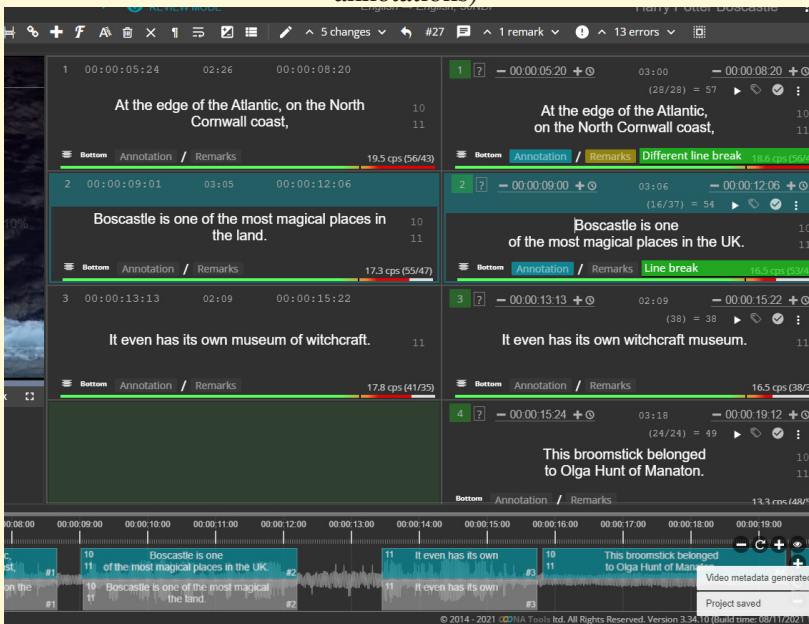


Source: Author

Participants received specific instructions, in the form of guidelines and other written materials, on producing and revising machine-generated subtitles. These had been automatically generated, from original audio in English and Spanish, employing Microsoft Stream's built-in ASR tool. The subtitle templates – i.e., the auto-timed intralingual transcriptions – were downloaded in SRT format and shared with the students via the project's dedicated file-sharing channel.

Student collaborators were responsible for the subsequent intralingual post-editing tasks. A smaller sub-team was responsible for the Spanish-to-English translation of the Spanish-language clip, for which the same ASR-generated subtitle template was obtained using Microsoft Stream. To undertake the subtitle post-editing tasks, students were given access to the professional version of cloud-based subtitling editor OOONA Tools (<https://oona.oonatools.tv>). They were trained on how to use the OOONA Review Pro editor. The interface of this tool, which can be seen in Figure 2 below, allowed the students to have a clear vision of what was being done in the editor.

**Figure 2:** Interface of OOONA Review Pro (including annotations)



Source: Author

Each student was allocated a mutually agreed, reasonable amount of editing work. All modifications committed in the post-edited templates were automatically logged by the tool, and participants could also make annotations to share with their counterparts for revision purposes. They could also generate a summary of revisions and compare both file versions at a glance. All the post-edited subtitle files were peer-reviewed by fellow participants, who were advised to aim for the highest quality possible (i.e., *publishable* quality). Therefore, participants were advised to undertake a *full* post-editing of the templates, followed by a detailed review of the resulting templates. Participants could discuss any challenging sections and share their revisions, as well as request feedback from staff members through the dedicated communication channel. Meanwhile, the 60-minute Spanish-language subtitle template, also produced using Microsoft Stream, was subsequently post-edited by two members of the Spanish-speaking team and translated into English by two English-to-Spanish trainees who cross-reviewed their work. Upon completion of their revision, participants exported the post-edited subtitle file. After the peer-review of the subtitle files was completed, staff members carefully completed quality control checks and produced the final templates. Staff members also provided feedback to the relevant post-editors and reviewers before liaising with the end-client to produce the final versions of the video chapters. The final subtitles were burnt in the videos using the OOONA Burn & Encode tool using accessible parameters such as non-serif and readable fonts and ultimately sent to the end-client to request feedback.

This situated learning experience focused on teaching ASR-generated subtitle post-editing in a real-world media localisation project that used authentic materials for publication. To foster their editing skills, participants carried out a series of quality checks whose ultimate aim was enhancing technical accuracy as well as the linguistic correctness of the ASR-generated subtitles. A post-project questionnaire, containing an informed consent form, was distributed among the participants. The responses obtained were extracted and analysed using Microsoft's online survey tool (MS

Forms). Due to the low number of responses and the nature of this situated learning experience, the discussion of the results in the next section centres on the qualitative value of the responses and capitalises on how this type of experience can inform the teaching of ASR followed by post-editing in subtitler training environments.

## **Results and discussion**

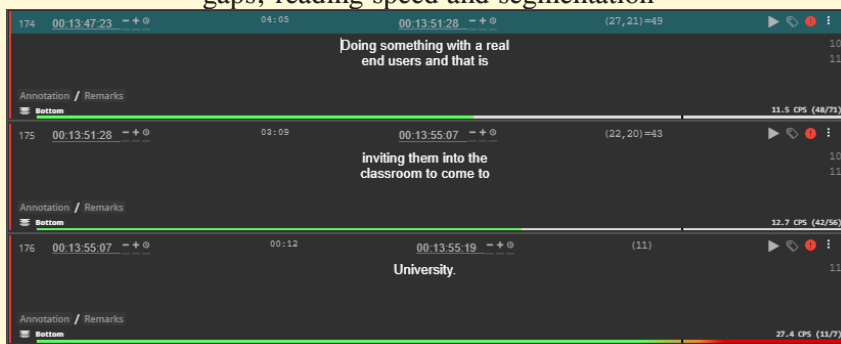
Among the participants (N=8), five filled in the questionnaire at the end of the project. Participants were either postgraduate students from the UCL's Centre for Translation Studies or recent graduates (or postgraduate students) from the University of Córdoba, Spain. They were all under 30 years of age and had completed undergraduate studies in modern foreign languages or translation. All of them had previously received subtitling-specific training at both undergraduate and postgraduate levels, though one of them reported having acquired subtitling skills through independent courses alongside their undergraduate course. As for post-editing, only one of them had received no training at all, which could have been considered a limitation to the experience insofar as their training is concerned. Needless to say, insufficient post-editing training can lead to situations in which post-editors over-edit and have an uninformed perception of post-editing; therefore, the *ad hoc* training and materials provided to the students aimed to attenuate the negative impact of this situation. Although still undertaking postgraduate studies at university, some of them were working as translators or subtitlers on a regular basis, and two of them had between one and three years of experience in the industry (which is an interesting datum given their early age, as mentioned above).

The participants' perception of the quality of the ASR-generated output was gauged to establish the perceived accuracy of the transcription on a scale from 0 to 10 (the latter being the highest quality possible). The technical dimension encompassed the spotting, formatting and layout of the subtitle lines, whereas

the linguistic dimension referred to word recognition as well as grammar and spelling.

On the one hand, the average rating of the technical quality of the spotting was 4.40 out of 10 (three respondents chose 5 while the other two chose 3 and 4, respectively). The most common problems reported by the students affected timecodes and synchronisation (including the automatic enforcement of minimum gaps) as well as segmentation and line breaks. Figure 3 below shows randomly selected consecutive subtitles that do not have a minimum separation gap – i.e., the end time of a subtitle coincides with the start time of the one that follows, but there should be a couple of frames to separate subtitles that appear one after the other. Additionally, one of the subtitles (comprised of one single word) has such a short duration (12 frames instead of the usual 1 second) that the resulting display rate (27.4 characters per second) would make it virtually imperceptible for the average viewer. These errors, however, can be easily amended by carrying out a technical check after setting up the file properties using a professional subtitle editor.

**Figure 3:** Subtitle editor showing issues regarding minimum gaps, reading speed and segmentation



Source: Author

The average rating of the linguistic quality, on the other hand, was considerably higher (6.20 out of 10), though not entirely satisfactory, with four respondents choosing 7 and one choosing 3.

The latter did not offer much detail on why their score was on the lower end of the spectrum; however, the rest of the respondents commented positively on punctuation and the fact that fairly few corrections had to be made. Misrecognised words were mainly caused by the presence of geographical names (e.g., “Cardbus” instead of “Córdoba”, “in crack off” instead of “Krakow”), people’s names (“John Maria” instead of “Gian Maria”) and cryptic acronyms (e.g., “as DH” instead of “SDH”), and so forth. The accuracy of automatic transcriptions is dependent on the audio quality of the source video and the clarity of the diction as well as how accented speech is processed, not least because foreign or non-neutral accents can be difficult to recognise for ASR engines (Kitashov, Svitanko & Dutta, 2018). Indeed, one of the participants gave a rather insightful reflection on the limitations of ASR tools: “Some of our speakers are not English natives, so the different accents may interfere. And when they gave an online talk, they might become nervous and hesitate when they speak English. [...] the shaking internet connections might have a negative influence on the sound quality and the speech recognition.” These factors are particularly relevant for educational materials inasmuch as conference talks and recorded lectures often take place live, where there is plenty of room for improvisation. Educational videos can feature a variety of accents as well as other elements such as room noise and interruptions, which may detract from the quality of the delivery and ultimately the audio track.

Respondents heavily criticised the many instances of poor segmentation and awkward sentence division that were present in the subtitles. The ASR tool transcribed text without considering subtitling-specific conventions such as line-breaking and characters-per-line rules<sup>2</sup>, as illustrated by Figure 4 below, which shows how the original subtitles did not follow conventional line-breaking

---

<sup>2</sup> As seen in similar studies, such as Tardel (2020) and Vitikainen & Koponen (2021), most ASR tools are not designed to comply with subtitling-specific guidelines. However, there are paid professional platforms, such as Matesub (<https://matesub.com>), which allow users to decide on gaps, number of lines and other subtitling-specific conventions.

principles and split up structures that should always appear together (e.g., “lots of people”, “are you (...)”). Participants claimed that the segmentation of the dialogue was not conducive to a comfortable reading of the subtitles either, and they, therefore, needed to merge and remove many subtitles (see crossed-out red box on the left-hand side of Figure 4).

**Figure 4:** Subtitle editor with the raw subtitle template and the post-edited template



Source: Author

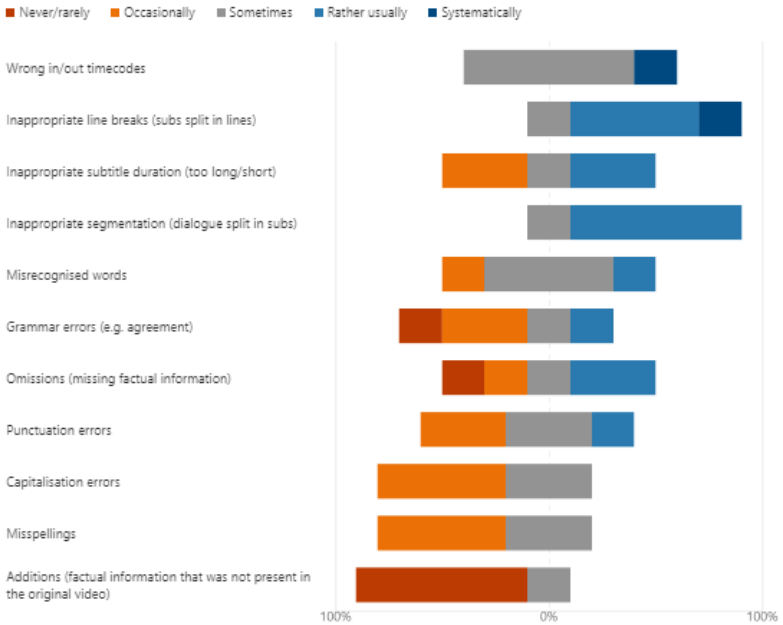
Respondents had to give their opinion on the most frequent errors that had been identified while post-editing the machine-generated subtitle. According to the results found in Figure 5, the most frequent errors were inaccurate timing, inappropriate line breaks and poor segmentation (i.e., technical accuracy). The least frequent errors were additions, misspellings, and capitalisation (i.e., linguistic correction). In the open-ended questions, respondents reported an infelicitous number of omissions, misrecognitions, and duration issues. They also expressed concerns about language correctness (grammatical and lexical infelicities). The different challenges experienced by the respondents arguably depend on the video files with which they worked and how the ASR tool performed in each instance. Other reasons include the participants’ lack of experience in ASR technologies and subtitle post-editing, but also the fact that they might have focused on the above-mentioned low technical quality.



**Figure 5:** Frequent errors in machine-generated subtitle templates

17. When post-editing the templates, what were the most frequent errors?

[More Details](#)

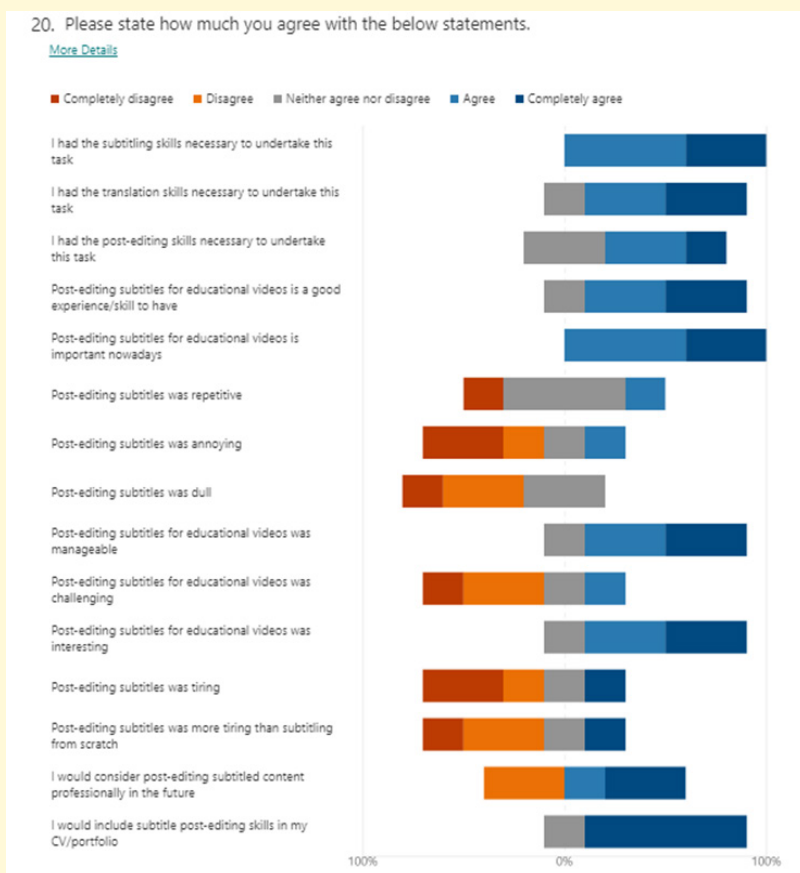


Source: Author

The above results indicate that linguistic errors were perceived to be far less present in the ASR-generated templates, with virtually no additions being made and grammar being sufficiently good, in their opinion. The widespread presence of subtitling-specific technical errors, such as timecodes and duration, alongside the many instances of poor segmentation and line breaks, show that commonly used ASR tools have not been designed to produce subtitles of a high professional standard. Most off-the-shelf ASR tools provide timed transcriptions unsuitable for publication or distribution in formal settings.

Students were then asked about post-editing from different angles. Indeed, their perception of their translation, subtitling and post-editing skills was gauged alongside other elements, such as the nature of the tasks and the usefulness of post-editing to localise educational videos. Interestingly, the closed questions indicate that the students seemed rather positive about post-editing overall (see Figure 6).

**Figure 6:** Participants' opinions on subtitle post-editing



Source: Author

The questions on whether they were ready to post-edit subtitled material and whether they considered it a good experience and important for educational videos were very positively rated. Only one student considered subtitle post-editing to be repetitive, annoying or challenging, and again only one of them would have preferred to produce the subtitles from scratch. This is consistent with previous studies in which students were exposed to similar machine-generated content, e.g., Matamala *et al.* (2015) and Moorkens (2018), among others. Perhaps one of the most intriguing results is the fact that, although all five respondents considered subtitle post-editing interesting, manageable and a good skill to have, two of them would not like to do it professionally. On the understanding that post-editing is increasingly present in the translation market (not only in AVT), four would also include it as a technical skill in their CVs or professional portfolios. This may indicate that participants were not entirely satisfied with the post-editing tasks – at least in their current form and because of the many errors caused by current ASR engines – while showing they were conscious of the manifold effects that artificial intelligence and automation technologies have on the translation profession. These results should be compared to similar studies in which reflections revolve around pay, experience and cognition (e.g., Da Silva & Costa, 2020), but also those that posit that post-editing training is becoming more normalised these days (Guerberof-Arenas & Moorkens, 2019).

In the open-ended question, participants could offer more detailed responses, which include statements in support of subtitle post-editing as a new, relevant practice, such as the following: “Subtitling a video from scratch might be a bit overwhelming”; “while computers will never be able to do as good a job as we can (thankfully!), they can help us to work more effectively and efficiently”; “subtitle post-editing is getting more popular and people tend to perform this type of task more often”; and “we can’t deny reality and we have to adapt our skills to the new labour work.” Vitikainen & Koponen’s (2021) argued that machine-generated,

auto-timed subtitles were produced more slowly on average than respeaking, and it seems that the participants from this small-scale study had a similar experience. These results, however, would have to be further evaluated, taking into consideration that more training on subtitle post-editing would be necessary since many “[...] students tend to think that any edit is valid as they are accustomed to editing their own work” (Guerberof-Arenas & Moorkens, 2019, p. 224), especially when considering they have little post-editing experience, as the questionnaire responses suggest.

Overall, respondents considered this experience to be useful but seemed equally aware that ASR technologies have many shortcomings when used in professional subtitling. This echoes Tardel’s (2020, p. 99) study, which concluded that “[...] current state-of-the-art ASR seems not to come near the effects correct transcripts have in terms of temporal, technical and cognitive effort.” Several participants would be keen on including post-editing as part of their hard skills when looking for jobs in the industry (“saying that subtitle post-editing does not enhance employability would be going against the trend”), but they would not necessarily enjoy it professionally. Arguably, this paradoxical situation experienced by would-be subtitlers is a reality that is here to stay as automation tools evolve and translation curricula struggle to embed industry-led transformations in the classroom. In light of the comments made by participants concerning their perception of the task, it would be worthwhile adding cognition to the equation and examining the role played by *cognitive effort* (Krings, 2001) in subtitle post-editing following research methods such as those proposed by Koponen *et al.* (2020) and Tardel (2021). Further research is needed to ascertain the extent to which automation tools, such as ASR engines, can support the translation and editing processes in subtitling.

## **Conclusion and final remarks**

This article has examined the post-editing of ASR-generated content in subtitler training environments. The discussion revolves around a situated learning experience involving subtitling trainees (mainly postgraduate students) who post-edited machine-generated subtitle templates for an accessible academic video book on media accessibility. Despite the rising interest in MT among industry stakeholders, as well as the subsequent responses from professional translators associations (Audiovisual Translators Europe, 2021), the use of ASR to generate subtitle templates has received scarcer attention from AVT scholars. Interest in ASR technologies has traditionally led to scholarly contributions in respeaking and live captioning rather than post-synchronisation subtitling or their use in subtitler training programmes.

In light of the swift technological changes led by industry developments, the pedagogical uses and applications of ASR tools in the subtitling classroom are worth being further explored. As a case in point, this learning experience prompted participants to undertake authentic revision tasks involving the editing of machine-generated subtitle templates. This learning experience prompted voluntary participants to edit ASR-generated content for publication purposes. Despite the fact that this authentic, situated project was positively rated overall, the respondents raised major issues concerning the quality of the machine-generated subtitle templates. Participants were particularly concerned about the poor technical quality of the templates and the ASR tools' inability to apply subtitling-specific conventions. The main issues affecting the ASR-generated subtitles used in this study had to do not only with linguistic aspects such as word recognition (particularly specialised terms and proper names) but, more importantly, with the fact that subtitling conventions (e.g., segmentation, line breaks, respect of shot changes) had not been observed. As a result, the subtitle templates that were automatically generated with Microsoft Stream, an off-the-shelf video-streaming platform with a built-in

ASR engine, had considerable room for improvement and could not be used without undertaking *full* post-editing of the content.

The identification and amendment of errors using professional subtitling systems constitute a legitimate approach to the learning and teaching of subtitling conventions, especially in terms of adjusting text-timed output to specific guidelines and technical parameters. The feedback received, in the form of a post-project questionnaire, confirmed this assumption. Participants reported on the benefits of reviewing ASR output as well as performing checks on each other's revisions. The positive feedback on this particular learning experience indicates that the use of automatic transcriptions in subtitle format can be helpful for students to spot text-timing errors and amend them by applying subtitling conventions seen in the classroom. Moreover, most students considered that subtitle post-editing was important and useful for disseminating educational videos. A paradox was identified inasmuch as the participants were aware of the importance of post-editing in today's industry but expressed a reluctance to undertake post-editing projects (at least willingly) in future professional endeavours. In future, a human-computer interaction approach to the teaching of ASR-produced subtitle templates could be envisaged. The fact that students handle ASR tools themselves (preceded and followed by editing as and when necessary) could lead to an improved perception of the task for use in the professional realm.

In the last decade, Macklovitch (2015, p. 575–576) stated that “[...] the fact that the [ASR] technology is not yet satisfactorily integrated with the other support tools that translators have come to rely on”, but ASR technologies have nowadays been democratised (e.g., YouTube Studio) and are also being further embedded in commercial subtitling software for professional purposes (e.g., MateSub and OOOONA). With the rise of ASR tools in the subtitling industry, however, it only follows that future subtitlers will be further exposed to scenarios in which template revision, editing and quality checks feature alongside more traditional approaches such as timing text from scratch or non-assisted template origination.

It is important that future AVT specialists recognise poor-quality machine-generated subtitles to decline jobs that would be too time-consuming or not worth the effort considering pay and working conditions. If exposed to real-world subtitling projects at an early stage, subtitlers-to-be can identify the main challenges they may encounter later in their careers, particularly regarding post-editing work, thus allowing them to make informed decisions when negotiating with agencies and end-clients. Following this principle, this article has shed light on the teaching of post-editing of ASR-generated content, showcasing a project-based, situated learning experience that exposed postgraduate students and recent graduates to localising authentic materials for publishing an accessible video book. It is hoped that this initiative can help foster further international collaborations of situated AVT learning experiences as well as raise awareness about the importance of accessibility when publishing educational materials in video format.

### **Acknowledgements**

This research was initially funded by UCL ChangeMakers (PGR Grants, 2021–2022). It took the form of an international collaboration between UCL’s Centre for Translation Studies, United Kingdom, and the University of Córdoba, Spain. The research has now been published thanks to funding from the European Union – Next Generation EU (“Margarita Salas”) from Universitat Jaume I (grant holder MGS/2022/03), Spain, and research group TRAMA. Many thanks to Dr Azahara Veroz-González and Dr María del Mar Ogea-Pozo (University of Córdoba, Spain) for participating in this educational experience with their students.

## References

Allen, Jeff. “Post-editing”. In: Somers, Harold (Ed.). *Computers and Translation: A Translator’s Guide*. Amsterdam: John Benjamins, 2003, p. 297-317.

Athanasidi, Rafaella. *The Applications of Machine Translation and Translation Memory Tools in Audiovisual Translation: A New Era?* Dissertation (Master of Science in Specialised Translation). Centre for Translation Studies, University College London, London, 2015.

Athanasidi, Rafaella. “The Potential of Machine Translation and Other Language Assistive Tools in Subtitling: A New Era?”. In: Bogucki, Łukasz (Ed.). *Audiovisual Translation – Research and Use*. Bern: Peter Lang, 2017. p. 29-49.

Audiovisual Translators Europe. *AVTE Machine Translation Manifesto*. Paris: Audiovisual Translators Europe, 2021. Available at: <https://avteurope.eu/avte-machine-translation-manifesto/>. Accessed on: May 7, 2023.

Bogucki, Łukasz & Deckert, Mikołaj (Eds.). *The Palgrave Handbook of Audiovisual Translation and Media Accessibility*. London: Palgrave Macmillan, 2020.

Bolaños-García-Escribano, Alejandro & Declercq, Christophe. “Editing in Audiovisual Translation (Subtitling)”. In: Sin-wai, Chan (Ed.). *Routledge Encyclopedia of translation Technology*. London: Routledge, 2023. p. 565-581.

Bolaños-García-Escribano, Alejandro & Díaz-Cintas, Jorge. “The Cloud Turn in Audiovisual Translation”. In: Bogucki, Łukasz & Deckert, Mikołaj (Eds.). *The Palgrave Handbook of Audiovisual Translation and Media Accessibility*. London: Palgrave, 2020. p. 519-544. DOI: [https://doi.org/10.1007/978-3-030-42105-2\\_26](https://doi.org/10.1007/978-3-030-42105-2_26)

Bolaños-García-Escribano, Alejandro; Díaz-Cintas, Jorge & Massidda, Serenella. “Latest Advancements in Audiovisual Translation Education”. *The Interpreter and Translator Trainer*, 15(1), p. 1-12, 2021. DOI: <https://doi.org/10.1080/1750399X.2021.1880308>



Bolaños-García-Escribano, Alejandro; Veroz-González, Azahara & Ogea-Pozo, María del Mar (Eds). *Media Accessibility in Modern Languages and Translation*. Córdoba: UCO Press, 2021.

Bowker, Lynne. *Computer-aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press, 2002.

Bowker, Lynne. “Computer-Aided Translation: Translator Training”. In: Sin-wai, Chan (Ed.). *The Routledge Encyclopedia of Translation Technology*. London: Routledge, 2015. p. 88-104.

BSI. *Translation services - Post-editing of machine translation - Requirements*. London: British Standards Institute, 2015.

Burchardt, Aljoscha; Lommel, Arle; Bywood, Lindsay; Harris, Kim & Maja Popovićet. “Machine Translation Quality in an Audiovisual Context”. *Target*, 28(2), p. 206-221, 2016. DOI: <https://doi.org/10.1075/target.28.2.03bur>

Bywood, Lindsay; Georgakopoulou, Panayota & Etchegoyhen, Thierry “Embracing the Threat: Machine Translation as a Solution for Subtitling Embracing the Threat: Machine Translation as a Solution for Subtitling”. *Perspectives*, 25(3), p. 492-508, 2017. DOI: <https://doi.org/10.1080/0907676X.2017.1291695>

Da Silva, Igor Antonio Lourenço & Costa, Cynthia Beatrice. “A tradução automática de textos literários e o comportamento do tradutor em formação: reflexões para o ensino da tradução”. In: Sousa, Germana Henriques Pereira; Costa, Patrícia Rodrigues & D’Ávila, Rodrigo (Eds.). *Formação de tradutores: desafios da sala de aula*. Campinas: Pontes Editores, 2020. p. 143-166.

Defrancq, Bart & Fantinuoli, Claudio. “Automatic Speech Recognition in the Booth”. *Target*, 33(1), p. 73-102, 2021. DOI: <https://doi.org/10.1075/target.19166.def>

Delabastita, Dirk. “Translation and Mass-Communication: Film and T.V. Translation as Evidence of Cultural Dynamics”. *Babel*, 35(4), p. 193-218, 1989. DOI: <https://doi.org/10.1075/babel.35.4.02del>

Díaz-Cintas, Jorge. “Clearing the Smoke to See the Screen: Ideological Manipulation in Audiovisual Translation”. *Meta*, 57(2), p. 279-293, 2012. DOI: <https://doi.org/10.7202/1013945ar>

Díaz-Cintas, Jorge. “Technological Strides in Subtitling”. In: Sin-wai, Chan (Ed.). *The Routledge Encyclopedia of Translation Technology*. London: Routledge, 2023. p. 720-731.

Díaz-Cintas, Jorge & Massidda, Serenella. “Technological Advances in Audiovisual Translation”. In: O’Hagan, Minako (Ed.). *The Routledge Handbook of Translation and Technology*. London: Routledge, 2019. p. 255-270.

Gambier, Yves & Gottlieb, Henrik. “Multimedia, Multilingua: Multiple Challenges”. In: Gambier, Yves & Gottlieb, Henrik (Eds.). *(Multi)media Translation*. Amsterdam: John Benjamins, 2001. p. 8-20. DOI: <https://doi.org/10.1075/btl.34.01gam>

Georgakopoulou, Panayota. “Technologization of Audiovisual Translation”. In: Pérez-González, Luis (Ed.). *The Routledge Handbook of Audiovisual Translation*. London: Routledge, 2018. p. 516-539, DOI: <https://doi.org/10.4324/9781315717166-32>

Georgakopoulou, Panayota. “Template Files: The Holy Grail of Subtitling”. *Journal of Audiovisual Translation*, 2(2), p. 137-160, 2019. DOI: [10.47476/jat.v2i2.84](https://doi.org/10.47476/jat.v2i2.84)

Georgakopoulou, Panayota & Bywood, Lindsay. “MT in Subtitling and the Rising Profile of the Post-Editor”. *Multilingual*, 25(1), p. 24-28, 2014. Available at: <https://multilingual.com/articles/mt-in-subtitling-and-the-rising-profile-of-the-post-editor>. Accessed on: May 7, 2023.

González-Davies, Maria & Enríquez-Ráido, Vanessa. “Situated Learning in Translator and Interpreter Training: Bridging Research and Good Practice”. *The Interpreter and Translator Trainer*, 10(1), p. 1-11, 2016. DOI: <https://doi.org/10.1080/1750399X.2016.1154339>

Guerberof-Arenas, Ana & Moorkens, Joss. “Machine Translation and Post-editing Training as part of a Master’s Programme”. *The Journal of Specialised Translation*, 31, p. 217-238, 2019.

Hickey, Sarah. “The 2022 NIMDZI 100: The Ranking of the Top 100 Largest Language Service Providers”. 21/02/2023. *Nimdzi*. Available at: <https://www.nimdzi.com/nimdzi-100-top-lsp>. Accessed on: May 7, 2023.

Hu, Ke & Cadwell, Patrick. “A Comparative Study of Post-editing Guidelines”. *Baltic Journal of Modern Computing*, 4(2), p. 346-353, 2016.

Kitashov, Fedor; Svitanko, Elizaveta & Dutta, Debojyoti. “Foreign English Accent Adjustment by Learning Phonetic Patterns”. *ArXiv:1807.03625*, p. 1-5, 2018. DOI: <https://doi.org/10.48550/arXiv.1807.03625>

Koponen, Maarit; Sulubacak, Umut; Vitikainen, Kaisa & Tiedemann, Jörg. “MT for Subtitling: Investigating Professional Translators’ User Experience and Feedback”. *Proceedings of the 14<sup>th</sup> Conference of the Association for Machine Translation in the Americas*, 14., 2020. p. 79-92. Available at: <https://aclanthology.org/2020.amta-research.pdf>. Accessed on May 7, 2023.

Kordoni, Valia; Birch, Lexi; Buliga, Ioana; Cholakov, Kostadin; Egg, Markus; Gaspari, Federico; Georgakopoulou, Yota; Gialama, Maria; Hendrickx, Iris; Jermol, Mitja; Kermandis, Katia; Moorkens, Joss; Orlic, Davor; Papadopoulos, Michael; Popovic, Maja; Sennrich, Rico; Sosoni, Vilemini; Tsumakos, Dimitrios; van den Bosch, Antal; van Zaanen, Menno & Way, Andy. “TraMOOC (Translation for Massive Open Online Courses): Providing reliable MT for MOOCs”. *Baltic Journal of Modern Computing*, 4(2), p. 217, 2016.

Krings, Hans. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Edited and translated by Geoffrey S. Koby. Kent: Kent State University Press, 2001.

Läubli, Samuel & Orrego-Carmona, David. “When Google Translate Is Better than Some Human Colleagues, Those People Are No Longer Colleagues”. *Translating and the Computer*, p. 59-69, 2017. DOI: <https://doi.org/10.5167/uzh-147260>

Lobato, Ramon. “Rethinking International TV Flows Research in the Age of Netflix”. *Television & New Media*, 19(3), p. 241–256, 2018. DOI: <https://doi.org/10.1177/1527476417708245>

Macklovitch, Elliott. “Translation Technology in Canada”. In: Sin-wai, Chan (Ed.). *The Routledge Encyclopedia of Translation Technology*. London: Routledge, 2015. p. 267-278.

Matamala, Anna. “The ALST project: Technologies for Audiovisual Translation”. In: *Proceedings of Translating and the Computer 37, 26–27 November, London, UK*. Geneva: Tradulex, 2015. p. 79-89. Available at: <http://www.tradulex.com/varia/TC37-london2015.pdf>. Accessed on: May 7, 2023.

Matamala, Anna. “Mapping Audiovisual Translation Investigations: Research Approaches and the Role of Technology”. In: Bogucki, Lukasz (Ed.). *Audiovisual translation – Research and Use*. Bern: Peter Lang, 2017, p. 11-28.

Matamala, Anna; Oliver, Andreu; Álvarez, Aitor & Azpeitia, Andoni. “The Reception of Intralingual and Interlingual Automatic Subtitling: An Exploratory Study within the HB4ALL Project”. In: *Proceedings of Translating and the Computer 37, 26–27 November, London, UK*. Geneva: Tradulex, 2015. p. 12-17. Available at: <http://www.tradulex.com/varia/TC37-london2015.pdf>. Accessed on: May 7, 2023.

Mehta, Sneha; Azarnoush, Bahareh; Chen, Boris; Saluja, Avneesh; Misra, Vinith; Bihani, Ballav & Kumar, Ritwik. “Simplify-Then-Translate: Automatic Preprocessing for Black-Box Machine Translation”. *ArXiv:2005.11197*, p. 1-8, 2020. DOI: <https://doi.org/10.48550/arXiv.2005.11197>

Mejías Climent, Laura & de los Reyes Lozano, Julio. “Traducción automática y posesición en el aula de doblaje: resultados de una experiencia docente”. *Hikma*, 20(2), p. 203-227, 2021. DOI: <https://doi.org/10.21071/hikma.v20i2.13383>

Moorkens, Joss. “What to Expect from Neural Machine Translation: A Practical In-class Translation Evaluation Exercise”. *The Interpreter and Translator Trainer*, 12(4), p. 375-387, 2018. DOI: <https://doi.org/10.1080/1750399X.2018.1501639>

Moorkens, Joss; O’Brien, Sharon; Da Silva, Igor Antonio Lourenço; Fonseca, Norma Barbosa de Lima & Alves, Fabio. “Correlations of Perceived Post-Editing Effort with Measurements of Actual Effort”. *Machine Translation*, 29(4), p. 267-284, 2015. DOI: <https://doi.org/10.1007/s10590-015-9175-2>

Mossop, Brian. *Revising and Editing for Translators*. London: Routledge, 2020.

Nikolić, Kristijan. “The Pros and Cons of Using Templates in Subtitling”. In: Baños, Rocío & Díaz-Cintas, Jorge (Eds.). *Audiovisual Translation in a Global Context. Mapping and Ever-Changing Landscape*. London: Palgrave Macmillan, 2015. p. 192-202.

Nikolić, Kristijan & Bywood, Lindsay. “Audiovisual Translation: The Road Ahead”. *Journal of Audiovisual Translation*, 4(1), p. 50-70, 2021. DOI: <https://doi.org/10.47476/jat.v4i1.2021>

Pérez-González, Luis. *Audiovisual Translation: Theories, Methods and Issues*. London: Routledge, 2014.

Pérez-González, Luis (Ed.). *Routledge Handbook of Audiovisual Translation*. London: Routledge, 2018.

Quah, Anne. *Translation and Technology*. London: Palgrave Macmillan, 2006.

Rico, Celia. “La formación de traductores en traducción automática”. *Tradumàtica*, 15, p. 75-96, 2017. DOI: <https://doi.org/10.5565/rev/tradumatica.200>

Romero-Fresco, Pablo. *Subtitling through Speech Recognition: Respeaking*. Manchester: St Jerome, 2011.

Schwab, Klaus. *The Fourth Industrial Revolution*. London: Penguin, 2017.

Sokoli, Stavroula. “Temas de investigación en traducción audiovisual: la definición del texto audiovisual”. In: Zabalbeascoa, Patrick; Santamaría, Laura & Chaume, Frederic (Eds.). *La Traducción audiovisual. investigación, enseñanza y profesión*. Granada: Comares, 2005. p. 177-185.

Tardel, Anke. “Effort in Semi-Automatized Subtitling Processes: Speech Recognition and Experience during Transcription”. *Journal of Audiovisual Translation*, 3(1), p. 79-102, 2020. DOI: <https://doi.org/10.47476/jat.v3i1.2020.131>

Tardel, Anke. “Measuring Effort in Subprocesses of Subtitling”. In: Carl, Michael (Ed.). *Explorations in Empirical Translation Process Research. Machine*

*Translation: Technologies and Applications*. Berlin: Springer, 2021. p. 81-110. DOI: [https://doi.org/10.1007/978-3-030-69777-8\\_4](https://doi.org/10.1007/978-3-030-69777-8_4)

TAUS. *Machine Translation Post-editing Guidelines*. 2016. Available at: <https://info.taus.net/mt-post-editing-guidelines>. Accessed on: May 7, 2023.

Torrejón, Enrique & Rico, Celia. “Skills and Profile of the New Role of the Translator as MT Post-Editor”. *Tradumàtica*, 10, p. 166-178, 2012. DOI: <https://doi.org/10.5565/rev/tradumatica.18>

Valuates Reports. “Captioning and Subtitling Solution Market Size to Reach USD 476.9 Million by 2028 at a CAGR of 7.7%”. 30/06/2022. *CISION PR Newswire*. Available at: <https://www.prnewswire.com/in/news-releases/captioning-and-subtitling-solution-market-size-to-reach-usd-476-9-million-by-2028-at-a-cagr-of-7-7-valuates-reports-875808599.html>. Accessed on: May 7, 2023.

Vitikainen, Kaisa & Koponen, Maarit. “Automation in the Intralingual Subtitling Process: Exploring Productivity and User Experience”. *Journal of Audiovisual Translation*, 4(3), p. 44-65, 2021. DOI: <https://doi.org/10.47476/jat.v4i3.2021.197>

Zabalbeascoa, Patrick. “La traducción de textos audiovisuales y la investigación traductológica”. In: Chaume, Frederic & Agost, Rosa (Eds.). *La traducción en los medios audiovisuales*. Castelló de la Plana: Publicacions Universitat Jaume I, 2001. p. 49-56.

Recebido: 21/02/2023

Aprovado em: 17/04/2023

Publicado em maio de 2023

---

Alejandro Bolaños García-Escribano. Castellón de la Plana, Community of Valencia, Spain. E-mail: [abolanos@uji.es](mailto:abolanos@uji.es). London, United Kingdom. E-mail: [a.bolanos@ucl.ac.uk](mailto:a.bolanos@ucl.ac.uk). <https://orcid.org/0000-0003-3005-2998>.