

Integrating post-editing into the subtitling classroom: what do subtitlers-to-be think?

Alejandro Bolaños García-Escribano

Universitat Jaume I

Spain

abolanos@uji.es

&

Centre for Translation Studies

University College London

United Kingdom

a.bolanos@ucl.ac.uk

<https://orcid.org/0000-0003-3005-2998>

Jorge Díaz-Cintas

Centre for Translation Studies

University College London

United Kingdom

j.diaz-cintas@ucl.ac.uk

<https://orcid.org/0000-0002-1058-5757>

Abstract

In today's professional landscapes, new technologies have altered media localization workflows as much as practitioners' workstations and habits. A more comprehensive integration of automation tools, including (neural) machine translation systems, has been ushered in by the proliferation of cloud ecosystems. In a further technological drive in the production of subtitle projects, systems now integrate automatic speech recognition and can machine translate subtitles from pre-spotted templates. The rise of post-editors in media localization, specifically in subtitling, has been a reality for some time now, triggering the need for up-to-date training methods and academic curricula. It is against this backdrop that this article seeks to examine the perception of post-editing among trainees in subtitling. A total of four teaching experiences, conceived as practical experiments in interlingual subtitle post-editing (English into Spanish), involving postgraduate students from both Spain and the United Kingdom, are described here. The sample comprised 36 master's-level students enrolled in translator training programmes that have a focus on audiovisual translation. A mixed-methods approach was adopted for this study; after each experience, the feedback collated through online questionnaires has proved paramount to understanding the participants' opinions about post-editing in the subtitling classroom. Interestingly, most of the respondents believe that subtitle post-editing training should feature more prominently in translation curricula even though they have voiced their reluctance to undertake post-editing work professionally.

Keywords: Audiovisual translation, subtitling, machine translation, post-editing, translator training

1. Introduction

Society today is characterized by an ever-growing daily consumption of audiovisual content led mainly by streaming platforms and the provision of information on the internet (Easton, 2021). New technologies have altered both industry workflows and translators' workstations and habits, gaining momentum in the media localization industry in particular (Díaz-Cintas & Massidda, 2019). In this industry, a more comprehensive integration of automation tools, such as translation memory (TM) tools, machine translation (MT) systems and automatic speech recognition (ASR) can be observed (Burchardt et al., 2016). Alongside the proliferation of audiovisual translation (AVT) platforms, dedicated MT engines are increasingly becoming a staple tool in cloud subtitling ecosystems such as OONA, Plint and ZOOSubs.

The high volumes and pressing deadlines encountered in the audiovisual industry emphasize the necessity of finding new ways to reduce the time spent on localization, and to this end the use of ASR, TM and MT seems promising. The rising demand for post-editing work in the media localization industry, specifically in subtitling, has been a reality for some time now (Bywood et al., 2017; Georgakopoulou & Bywood, 2014); and although seasoned subtitlers and newcomers are facing increasing volumes of post-editing tasks, scholarly attention has been scarce on this front (Bolaños García-Escribano & Declercq, 2023). The need to train subtitlers in the uses and applications of MT and post-editing is closely related not only to market demand but also to the evolution of automation in the rapidly developing age of artificial intelligence. Interest in the way post-editing is perceived has grown in recent years across the language industries (Rossi & Chevrot, 2019); this study, with its focus on subtitle post-editing, draws on previous scholarly attempts (González Pastor, 2021; Moorkens et al., 2018; Pérez Macías, 2020) to establish the ways in which future generations of AVT professionals perceive automation tools, particularly MT.

Following a mixed-methods approach, this article sets out to examine the integration of subtitle post-editing into translator training environments and to gauge would-be subtitlers' perceptions of subtitling. This research comprised four pedagogical experiences conceived as practical experiments and involving advanced-level English-to-Spanish subtitling students, most of whom had previously received non-specialist post-editing training as part of their studies. Each participant received a clip and a subtitle project file containing a bilingual subtitle template (English original and Spanish translation), but there were three different versions: two translations that had been produced by two MT engines and one translation authored by a postgraduate subtitling student. The authored translation, in essence a revision task, was used as a distractor. After each experiment, the participants were prompted to complete an online questionnaire. The feedback collated has proved decisive in enabling a better understanding of the participants' opinions on subtitle post-editing. Interestingly, most of the respondents believed that subtitle post-editing training should be embedded further in the AVT curriculum; however, they reported low levels of satisfaction when asked about undertaking professional post-editing work in the future. This suggested an overall negative perception of post-editing subtitles as an occupation.

2. Machine translation and post-editing in audiovisual translation

Audiovisual texts are a composite of sounds and images, encompassing four different types of sign – audio-verbal, audio-nonverbal, visual-verbal and visual-nonverbal – that produce meaning in a final complex communicative output (Delabastita, 1989). AVT practices, including subtitling, involve the localization of audiovisual media content while paying attention not only to spoken text but also to onscreen kinetics, body language, and gestures in addition to sounds and cinematographic information. Since the turn of the millennium, the AVT professions have undergone fundamental changes as the media industries have multiplied and diversified; this has led to considerable academic interest in the complex semiotic texture of audiovisual texts (Bogucki & Deckert, 2020; Gambier & Gottlieb, 2001; Pérez-González, 2018).

Subtitling has long been the preferred mode of AVT on the web, largely because it is the fastest and most inexpensive media localization practice (Díaz-Cintas & Remael, 2021). For this reason, it has traditionally attracted the attention and efforts of companies aiming to implement automation tools, which have then been progressively integrated into the industry workflows (Burchardt et al., 2016). As posited by Díaz-Cintas (2023), in the current mediascape of digitalizing communication, rendering it increasingly audiovisual, and increasing the use of the internet as a communications channel, subtitled materials are being used more frequently on social media, video games and streaming platforms. Of this trend, technology has been a prominent ally, facilitating the transfer of a source-language text into a target version while aiming to enhance productivity and cost-efficiency.

Among the most frequent translation technologies, MT engines have started to be used more frequently since the advent of statistical MT (SMT), and especially neural MT (NMT). Automation technologies, particularly MT engines, were latecomers in AVT practices (Díaz-Cintas & Massidda, 2019), with some functionalities being developed mostly for subtitling programs. More recently, though, dubbing and audio description have also benefited from new technological developments. Specialist software is used to deal with the technical challenges of subtitling: agile spotting, the visualization of sound and shot changes, an indication of the display rate, and the like. Regarding the application of language automation to subtitling workflows, the developments have been more modest (e.g., spell-checkers); and TM and MT have not been too prominent (Athanasidi, 2017) until recently, when the expansion of cloud-based platforms has facilitated their integration (Bolaños García-Escribano & Díaz-Cintas, 2020; Bolaños García-Escribano et al., 2021). Nowadays, the number of cloud-based AVT systems that integrate automation technologies is sizeable.

Many research efforts have been devoted to understanding why and how the quality of MT's raw output compares to that of human translations (Läubli & Orrego-Carmona, 2017). To reach acceptable quality standards, machine-translated raw output has to undergo post-editing, that is, human revision of MT output (British Standards Institute, 2015). Revision is understood to be the bilingual examination of a human-generated target text (i.e., a translation) in direct comparison with the source text (Declercq, 2023). The task of post-editing differs from that of revision in many aspects since the types of editing that need to be made to MT output depend both on “the expected level of final quality and on the type of errors produced [by the engine]” (Mossop, 2020, p. 216). Post-editing guidelines are often

Bolaños García-Escribano, A., & Díaz-Cintas, J. (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137

language- and company-specific (Allen, 2003), which leads to a general lack of homogeneity in the industry. Despite that, there seems to be consensus, as explained by Hu and Cadwell (2016), that there are two main types of post-editing: light and full. Whereas the former aims at the production of an understandable and usable target text, even if it is not linguistically or stylistically perfect, the latter pursues the stylistic and linguistic correctness that resembles human translation.

The productivity of translators is determined not only by the efficiency of the technologies employed – that is, the more efficient the technology, the less the amount of post-editing that is needed – but also by the time invested in making the target files fit for purpose. In this sense, the use of MT followed by human post-editing makes sense only if the degree of editing is low and translators can improve the quantity and quality of their output. When MT tools are used, the time and cognitive input it takes to amend the suggested translations is often referred to as ‘post-editing effort’ (Krings, 2001), and this has been the focus of much of the editing-related research to date (do Carmo et al., 2021; Gaspari et al., 2014; Koponen, 2012, 2016; O’Brien, 2005, 2011; Specia & Farzindar, 2010). Regarding post-editing, some scholars have reported a growth in translators’ productivity when they do post-editing compared to when they translate from scratch in certain scenarios (Federico et al., 2012; Guerberof Arenas, 2008; Plitt & Masselot, 2010).

The challenges of using MT engines in subtitling are manifold (Burchardt et al., 2016) and they have been the subject of study of several EU-funded research projects, such as MUSA,¹ eTITLE,² SUMAT,³ and EU-BRIDGE.⁴ ASR, in combination with MT, has been tested in initiatives such as TransLectures⁵ with the aim of localizing academic video content. More recently, projects such as MultiMT⁶ have looked into ways of supplementing SMT by going beyond written text and devising methods and algorithms which exploit the multimodal information that is transmitted through the images and the audio tracks of audiovisual productions. Automation is indeed making far-reaching inroads into subtitling, with many companies developing paid and proprietary MT systems for in-house work, and with new approaches that use artificial neural networks becoming the mainstream MT engines in the industry.

To date, only a few studies have explored the use of MT in subtitling workflows, and, as in the case of non-audiovisual texts, their findings suggest productivity gains overall (Bywood et al., 2017; de Sousa et al., 2011; Koponen et al., 2020a, 2020b; Matusov et al., 2019), thus fuelling the need for further research on MT applications in the AVT field. One of the possible reasons behind the scarcity of literature on the topic resides in the technical challenges typical of

¹ Multilingual Subtitling of Multimedia content, 2002–2004, <https://cordis.europa.eu/project/id/IST-2001-38299>

² European Multilingual Transcription and Subtitling Services for Digital Media Content, 2004–2005, <https://cordis.europa.eu/project/id/22160>

³ An Online Service for SUBtitling by MACHine Translation, 2011–2014, <https://cordis.europa.eu/project/id/270919>

⁴ 2012–2015, www.eu-bridge.eu

⁵ Transcription and Translation of Video Lectures, 2011–2014, <https://cordis.europa.eu/project/id/287755>

⁶ Multi-modal Context Modelling for Machine Translation, 2016–2021, <https://cordis.europa.eu/project/id/678017>

subtitling, challenges which the current NMT systems have not fully resolved. For instance, subtitlers often have to condense or omit information to meet the subtitle display rates established in the job guidelines. Therefore, even if the raw MT output is of high quality and does not need much post-editing because it is semantically correct and adequate for the context, subtitlers might very well still have to edit it down or even discard it completely because of technical considerations such as the maximum display rate or the number of characters allowed per subtitle line.

In fact, according to Bolaños García-Escribano and Declercq (2023), there are “at least six different types of editing that traditionally take place in the AVT industry – i.e., pre-editing, post-editing, revision, proofreading, QC and post-QC viewing” (p. 576). A *supra* form of editing, known as truncation, can occur at any point while producing subtitles and it relates to the partial or total condensation of information to comply with the spatio-temporal restrictions that characterize subtitling (Díaz-Cintas & Remael, 2021). When using MT, truncation often still remains to be done, the required levels of truncation being dependent on whether the MT output has been generated from a pre-timed subtitle template containing a verbatim transcription of the original dialogue (which can be produced by either a human subtitler or an ASR engine) or from a pre-timed pre-edited template in which parts of the original exchanges have already been condensed to comply with spatio-temporal considerations such as character-per-line and display-rate limitations. The production of a subtitle template and the technologies used in the workflow downstream have an impact on the level of editing that is required. When a subtitle template has been pre-edited and pre-timed, the work that translators, post-editors, or revisers have to undertake is considerably more focused on both the linguistic components and the adherence of the resulting output to the specificities and guidelines of each project.

Another aspect that characterizes AVT projects is creativity, which has received much attention from scholars (Romero-Fresco & Chaume, 2022) and should not be neglected when post-editing subtitles. In a study on translators’ perceptions of NMT in literary translation, Moorkens et al. (2018) posited that “a further limitation to the broader use of MT is that post-editing is a task disliked by many translators, who have complained that it constrains their work, allows limited opportunities for creativity” (p. 241). Likewise, AVT associations, such as the pan-European AVTE and Spain’s ATRAE, have produced manifestos and news releases advising members against the unfair and unethical use of post-editing:

We urge (video-on-demand) platforms to speak to language service providers so that they refrain from using machine translation followed by post-editing and trust human talent with the job instead (ATRAE, 2021, online; our translation).

Given the negativity surrounding this practice, the ultimate goal of these experiments was to gauge the perception of students who are being trained in subtitle post-editing and to ascertain whether they still (dis)like this practice after being exposed to it in a semi-authentic experience.

3. Research materials and methodology

3.1. Experiment setup

In order to better understand the students' perceptions of subtitle post-editing practice in the classroom, four experiments were conducted to collect quantitative and qualitative data from would-be subtitling professionals who were exposed to subtitle MT and post-editing, from English into Spanish, as part of their studies in higher education. For this, a three-hour task-based lesson was designed and taught at master's level at one Spanish university in two different years (2021, 2022), at a British university in 2022, and at another Spanish university in 2022. All the participants (N=36) were postgraduate translation students with Castilian Spanish as their native language; they had completed AVT- and subtitling-specific training as part of their studies. The experiments were designed in the form of teaching experiences and the participants in each cohort were prompted to revise a different subtitle template, as explained below.

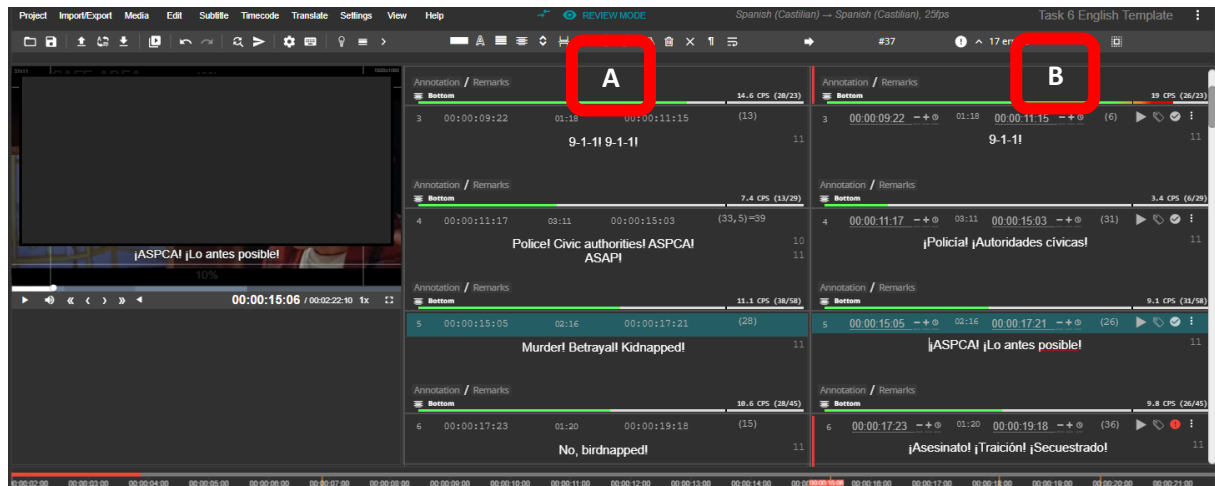
A 2 min 22 s scene from *Mrs Doubtfire* (Columbus, 1993) was used for the experiments together with the pre-timed template of the original English dialogue. The template, which was the same as that used in the commercial distribution of the film, consisted of a total of 37 subtitles, with 344 words and 1,829 characters with spaces. The template contained the full transcription of the dialogue, including repetitions and reformulations. Using a professional cloud-based subtitling system, namely OOONA Tools⁷, the template was translated into Castilian Spanish in two different ways, which produced three different working documents. In the first approach, a Spanish-speaking postgraduate student produced a translation (human template, HT), whereas in the second, the template was translated automatically (AT template) using two different non-domain-specific, built-in NMT engines: AppTek (AT1) and Amazon Translate (AT2). The HT was introduced into the experiments as a distractor with which to analyse the students' responses and to establish whether a potential negative bias against NMT post-editing could be traced. Both the HT and the AT templates were produced in 2021 before the first experiment took place.

The participants were informed of the existence of the three different Spanish templates and were randomly provided with one of them alongside the English template and the original video. They worked with OOONA Translate Pro, with which they were all familiar before the experiments began. The templates were distributed randomly, ensuring that more than three-quarters of the participants post-edited one of the two AT templates (AT1 or AT2). The participants had to take decisions regarding not only the translation but also the potential need for reduction in order not to transgress the maximum display rate of the Spanish subtitles, which had been set at 17 characters per second (cps), and the maximum number of characters per line, set at 42. Figure 1 displays the interface that the participants could see. The working area included the original English template in the left-hand column (A) and also the Spanish template that they had to revise in the right-hand column (B), where they could complete the revision and introduce changes and add comments:

⁷ See <https://edu.oonatools.tv>

Figure 1

Interface of the OONA Translate Pro tool



Before the experiments, the participants had been exposed to current post-editing practices in AVT. They had had lessons in subtitling workflows and had undertaken cross-revision exercises, so they were familiar with the principles of human-generated subtitle editing (including bilingual revision). The experimental task was preceded by a short lecture which focused on the basics of NMT and post-editing. O'Brien's (2011) definitions of critical, major and minor errors were fleshed out, and the participants were briefed on the nature and severity of the errors that are often found in subtitling, not only in post-editing but also when revising human-generated translations. They were handed a printed document containing a brief summary of each of the errors, grouped into three overarching categories (translation, linguistic and technical), in which they could keep a record of the number, nature, and severity of those they found in their template (Table 1):

Table 1

List of errors for subtitle revision and post-editing, inspired by O'Brien (2011)

Error		Critical	Major	Minor
TRANSLATION	ACCURACY (rendering of meaning and presence of sense-related issues)			
	CULTURAL TRANSFER (adaptation of cultural elements and humorous passages)			
	REGISTER, NATURALNESS AND IDIOMATICITY (appropriateness of the target text and adherence to target-language stylistic norms)			
	TOTAL:			
LINGUISTIC	CORRECTION (syntactical structures and vocabulary)			
	PUNCTUATION (adherence to orthotypographical norms)			
	CAPITALISATION (use of upper and lower case)			
	SPELLING (accuracy of word formation and spelling)			
	OMISSIONS (lack of information present in the video and needed to follow the plot)			

Bolaños García-Escribano, A., & Díaz-Cintas, J. (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137

	ADDITIONS (inclusion of information that is either present not present in the video but irrelevant to follow the plot)			
	TOTAL:			
TECHNICAL	DURATION (subtitles legibility in terms of display rate)			
	SEGMENTATION (dialogue division across subtitles)			
	LINE BREAKS (within subtitles)			
	TOTAL:			

For the subtitle post-editing task itself, the participants had to fix any errors they noticed in the Spanish template, for which they were given a set of instructions (Table 2):

Table 2
Instructions shared with participants

TRANSLATION	<ol style="list-style-type: none"> 1. Fix any meaning-related issues such as mistranslations (e.g., faux-sense, sans-sense) and semantic inaccuracies. 2. Fix any humorous or cultural references that cannot be easily understood by the target audience. 3. Fix any instances of inappropriate register, tone or degree of naturalness and orality.
LINGUISTIC	<ol style="list-style-type: none"> 4. Fix any terminological issues, either technical or non-technical, as well as any inconsistencies. 5. Fix grammatical and lexical infelicities, e.g., sentence structure, linear word order and idiomatic expressions. 6. Fix morphological errors related to plural and gender agreement, case, person, tense, etc. 7. Fix missing text (e.g., sentence, phrase, word) if the omission interferes with the message being transferred. 8. Fix any misspellings and typos. 9. Fix incorrect punctuation. 10. Do not fix (minor) stylistic issues unless they interfere with the message. 11. Fix any offensive or culturally unacceptable information in accordance with the creative intent of the original dialogue.
TECHNICAL	<ol style="list-style-type: none"> 12. Fix any subtitle lines that exceed the maximum number of 42 characters per line. 13. Fix any subtitle whose display rate is higher than 17 cps. 14. Fix any mismatch between the subtitles and the visuals that could be misleading for the viewer.

At the end of each session, an online survey containing 23 questions was distributed and 36 students completed it between 2021 and 2022. Using the data obtained and their statistical value, the discussion now focuses on understanding the participants' opinions of subtitle post-editing after having participated in a practical workshop.

Our working hypothesis was that the students would have a negative perception of subtitle post-editing despite it having become a frequent activity in today's industry and irrespective of the template with which they worked. We also anticipated that the AT templates might contain a greater number of errors of a more diverse nature, therefore making the task more onerous than simply revising a human-generated template and accentuating the participants' unfavourable perception of this practice.

3.2. Sample description

The sample of this study (N=36) that took part in the four experiments comprised 29 subjects who worked with the NMT-produced template translations (AT) and 7 who were assigned to the human-translated one (HT), both of the translations being in Castilian Spanish. All the participants were asked to supply some personal information such as their age, educational background, and professional experience.

Most of the respondents were in their early twenties (30; 83%), except for 4 of them, who were aged 25–30 (11%), and 2 who were older than 30 (6%). All of them had done subtitling as part of their university training, with 17 (48%) having completed at least a postgraduate module, 7 (19%) having done both undergraduate and the postgraduate modules, and 8 (22%) having completed an undergraduate module in subtitling. Only 4 (11%) of them had undergone extracurricular training. A total of 11 (30%) respondents claimed not to have been exposed to post-editing training during their studies, whereas the rest had covered post-editing during their undergraduate (10; 28%) or postgraduate studies (12; 33%), with a marginal number of the participants having learnt about post-editing across their university studies (1; 3%) or as an extracurricular activity (2; 6%).

All the respondents were native speakers of Castilian Spanish. Regarding their professional status, most of them (31; 86%) were still university students at the time the experiments took place and had no links to the industry, while 2 (6%) of them undertook the occasional freelance professional project and 3 (8%) were actively working in the language or translation industries alongside their studies. When asked about their experience in the translation industry, the majority (19; 53%) reported having had none whatsoever, while the remaining 17 (47%) claimed to have had some, of whom only 3 (8%) had had more than one year of experience. As can be expected from the answers to the previous question, 31 (86%) had no former post-editing experience in the industry, 5 (14%) had done training to some extent, and only 1 (3%) had had more than three years' experience in the field. All in all, it can be concluded that although many of the participants had some knowledge of the topic, this was more theoretical than practical.

4. Results and discussion

This section reports on the statistical data collected in the four experiments and their subsequent analysis in SPSS to enable the retrieval of significances and correlations. The data were collated from the various online questionnaires completed by the respondents on MS Forms.

4.1. Perception of (overall) subtitle quality

We sought to elicit the participants' opinions on the post-editing of machine-translated subtitles. But rather than comparing the quality of various NMT engines, the emphasis was placed on what the respondents made of the subtitle template that was randomly assigned to them. Indeed, they could only guess whether they were post-editing (AT group) or revising (HT group), and this helped us to ascertain whether pre-conceived ideas about post-editing would affect their answers.

To gauge their perception of the overall quality of the translated templates, the respondents were prompted to award a score out of 10 (1 being the lowest score and 10 being the highest) to the template given to them, taking into account both the translation and the technical and linguistic dimensions inherent in professional subtitling. The combined results of both groups reveal that the values are normally distributed, showing a (consistently) positive skewness of up to 1.265 for these questions. As illustrated in Table 3, the AT templates scored very low ($\mu=3-3.41$) whereas the HT's score was more than double ($\mu=6.71-7.29$); this indicated that the AT templates were perceived to be of a much poorer quality than the HT template. Having said that, the standard deviation values present in the HT group are particularly high ($\sigma > 3$ points for each category), meaning that some of the participants conferred much lower scores on these categories.

Table 3

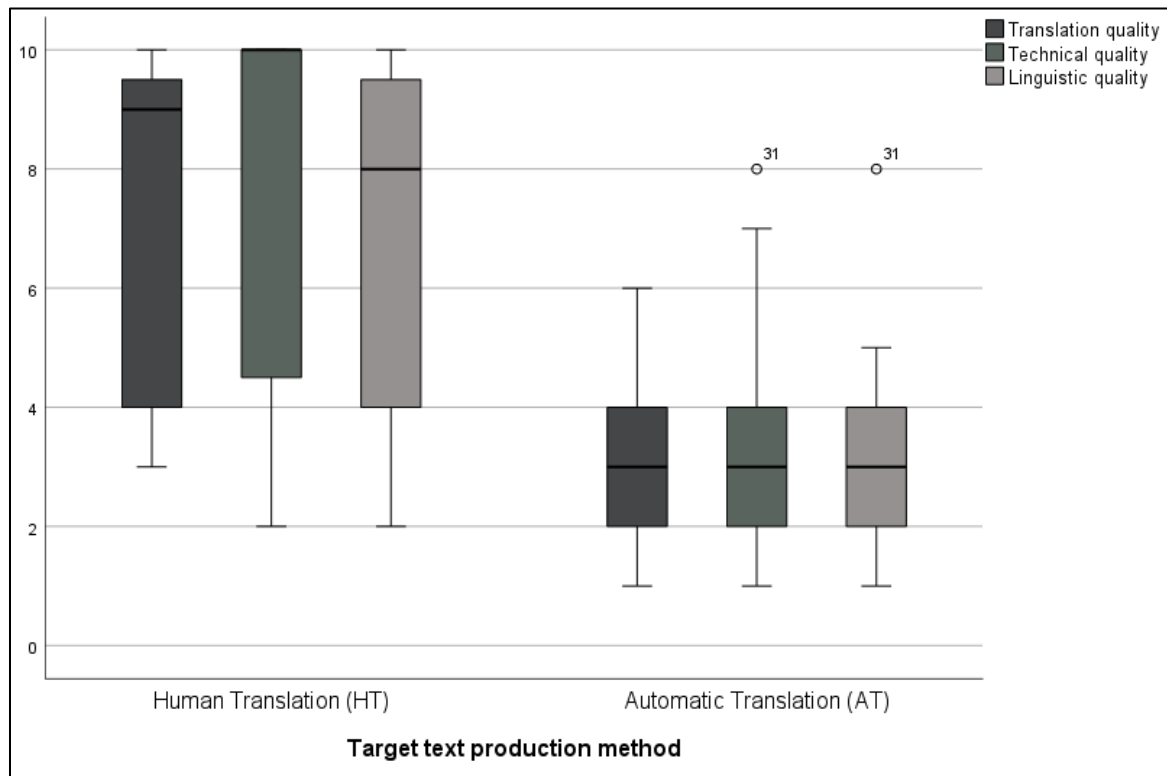
Frequency values for overall quality scores

N=29 (AT) N=7 (HT)	Mean μ		Median M		Std Deviation σ		Variance σ^2		Minimum		Maximum	
	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT
Translation quality	3.00	7.00	3	9	1.488	3.215	2.214	10.333	1	3	6	10
Technical quality	3.41	7.29	3	10	2.027	3.498	4.108	12.238	1	2	8	10
Linguistic quality	3.14	6.71	3	8	1.642	3.352	2.695	11.238	1	2	8	10

Since the variables show a Gaussian distribution, a Mann-Whitney U test ($\alpha=0.05$) was performed. The results show that there is a significant correlation between the text production method (whether AT or HT) and the score awarded to each quality category, with values being $p=.005$, $p=.009$ and $p=.012$ for the translation, technical and linguistic quality variables, respectively. The null hypothesis was that the distribution of the quality variables would be equal for groups AT and HT, but this assumption was discarded because the scores vary significantly depending on the group. Therefore, it can be concluded that the perceived quality of the HT is considerably higher than that of the AT. Figure 2 is a visual representation of the scores for both groups:

Figure 2

Overall translation scores in AT and HT groups



Except for one outlier, that is, a participant who rated the technical and linguistic quality of the AT templates much higher than the rest (8 out of 10), the median rating of all three categories for these templates was 3. The medians of the HT, however, were significantly higher: 9, 10 and 8 for translation, technical and linguistic quality, respectively. In the case of the AT, technical quality goes as high as 7 in the upper limit for the fourth quartile, although the upper limits for translation and linguistic quality were 6 and 5, respectively. The latter values are considerably lower than those of their HT counterparts, which are 10 in all three categories. The standard deviations affecting the HT can be seen in the lower limits for the first quartile (i.e., 3 for translation quality and 2 for technical and linguistic quality), which suggests that some of the participants were dissatisfied with the overall quality of the HT. The box plots are also considerably higher than expected, indicating obvious perception differences among HT responses compared to those for ATs, where there seems to be broader agreement on the perception that the overall quality of the templates is indeed low.

In the subsequent open-ended questions, the respondents were required to write a short reflection on their quality scores. The most frequent technical errors mentioned by the participants in both groups were these: deficient line breaks and segmentation, the appearance of three lines, poor legibility of the subtitles due to high display rates, and a lack of dashes to identify different characters within the same subtitle event. On the translation and linguistic fronts, the most frequent errors stated were these: incorrect syntax and lexis, meaning-related issues, lack of idiomaticity, literal transfer of connotations and jokes, wrong plural and gender agreements, punctuation conventions, literalness and calques, lack of cultural adaptation, stylistic mismatches, and unnaturalness. Interestingly, those participants

who reviewed the AT template made note of fewer errors in their responses to those questions.

The respondents from the HT group did not report any major technical issues, except for one of them, who argued that the template “didn’t follow any of the conventions”. Regarding to translation and linguistic quality, only four respondents from the HT group reported no major issues, whereas the remaining three were unhappy about the style, tone, and overall adaptation of the cultural and humorous content (one respondent stated that “the base language was correct, but the tone and the jokes were not translated right”). It is worth noting that this participant is the outlier discussed in the next section.

4.2. Identification of (specific) errors in templates

The results yielded by the answers to the quantitative and qualitative questions on the perception of the overall quality of the templates are telling in their own right. However, to understand the recurrence of errors in the Spanish templates better, the participants were prompted to rate from 1 (never or rarely) to 5 (systematically) the frequency of certain types of error – meaning that the higher the values analysed in this section, the poorer the perceived quality of the template being discussed. The data are Gaussian, showing a skewness of up to -0.412 and 0.513 , except for the omission and addition errors, which show greater skewness of 1.159 and 1.306 , respectively. This indicates that the respondents did not find too many omissions or additions in the templates, hence the positive inclination.

NMT engines do not modify the timecodes of templates to adhere more closely to the expected display rate; neither do they normally take into account grammar when breaking the lines. Subtitling systems, however, can be configured so that the target subtitles do not exceed the character-per-line limit when NMT output is retrieved. Therefore, the AT templates were expected to be both more verbatim and wordier and more poorly segmented than the HT, which could be modified by the template translator so as to align the text better with the maximum display rate of 17 cps.

As illustrated in Table 4, the mean of technical error frequency is significantly lower for the HT ($\mu < 2.5$ out of 5 for all five categories) than the AT templates ($\mu > 3.3$ and closer to 3.50 in the case of segmentation and line breaks):

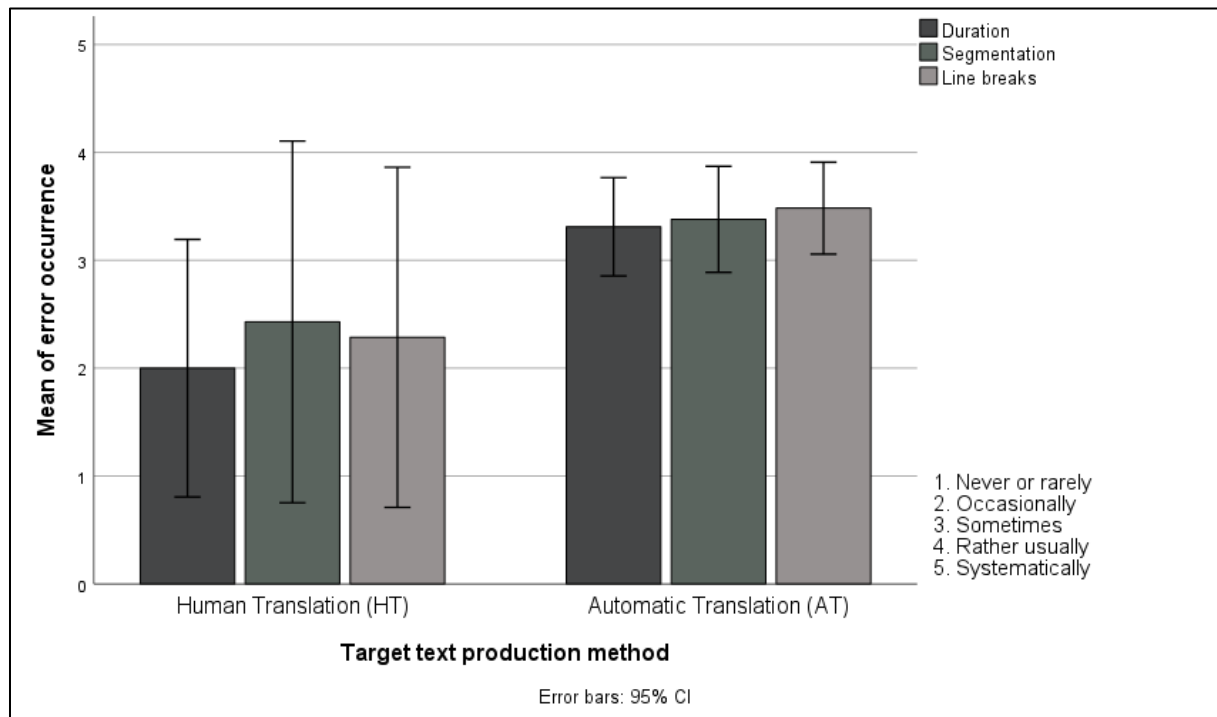
Table 4
Frequency values for technical errors

N=29 (AT) N=7 (HT)	Mean M		Median M		Std Deviation σ		Variance σ^2		Minimum		Maximum	
	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT
GROUP												
Duration	3.31	2.00	4	1	1.198	1.291	1.436	1.667	1	1	5	4
Segmentation	3.38	2.43	4	1	1.293	1.813	1.672	3.286	1	1	5	5
Line breaks	3.48	2.29	4	1	1.122	1.704	1.259	2.905	1	1	5	5

However, as shown in Figure 3, a more granular consideration of the results does not seem to support the tempting argument that the technical quality of the HT is substantially better than that of the AT templates:

Figure 3

Technical errors identified by the AT and HT groups



The application of a Mann-Whitney U test ($\alpha=0.05$) reveals that the duration variable can be seen as significant when the text production method ($p=0.22$) is considered. The spread of the highest value for the HT is particularly revealing for both segmentation and line breaks, with a 0.05 margin of error. This is so because the standard deviation is significant for the AT scores ($\sigma=1.813$ for segmentation; $\sigma=1.704$ for line breaks). It therefore becomes clear that the HT group was not fully satisfied with the technical dimension of their template. Such a deviation can be explained in various ways: the participants might have been particularly strict when revising the human-generated subtitles or they might have thought that they were editing an automatically produced template and unconsciously scored the template more negatively. Another assumption is that the outlier (i.e., Participant 3.14) marked the HT particularly harshly, thus affecting the upper quartiles. Be that as it may, and despite the fact that the overall values indicate that the HT was more positively perceived on the whole than the AT templates, the results do not necessarily suggest that the segmentation and line breaks in the AT templates were of a lesser quality.

Regarding the translation and linguistic dimensions, the participants rated the presence of syntactical and lexical infelicities (correction), unconventional punctuation, inappropriate capitalization, misspellings, omissions, and additions. These parameters enabled them to evaluate the adequacy of the target-language templates compared to the original dialogue (i.e., meaning transfer) and also to ascertain whether the subtitles were accurate and complete, whether any key information was missing, or whether any unnecessary wording had been introduced.

It is perhaps not surprising that language-related errors were more visibly present in the AT template than in the HT. According to the participants (Table 5), the AT was far from being

syntactically and lexically correct ($\mu=3.34$ out of 5), with little deviation ($\sigma=0.936$). This suggests that the two AT templates were not adequate linguistically speaking. Punctuation and capitalization obtained a higher score but still presented problems ($\mu=2.86$ and $\mu=2.45$, respectively), which may be due to the fact that NMT engines typically find it challenging to follow subtitling-specific punctuation conventions. Indeed, NMT developers often neglect these conventions, as has been observed in studies such as that conducted by Gupta et al. (2019). Remarkably, the AT template performed best at spelling ($\mu=1.69$ out of 5; $\sigma=0.967$) and was perceived to be more faithful than the HT to the English original dialogue contained in the template (i.e., omissions and additions never or rarely occurred, with $\mu=1.9$ and $\mu=1.62$, respectively). As it happens, the HT received higher scores overall, even if these are far from perfect. Indeed, despite showing much standard deviation ($\sigma=1.988$), the correctness of the HT is a healthy $\mu=2.57$. Just as in the previous sections, several HT participants did not consider the template to be entirely correct and commented on the document's containing errors related to translation and linguistic correction, punctuation, and omissions. The only error that shows a significantly better score is capitalization ($\mu=1.43$ and $\sigma=0.787$), with a maximum value of 3 out of 5 for the HT.

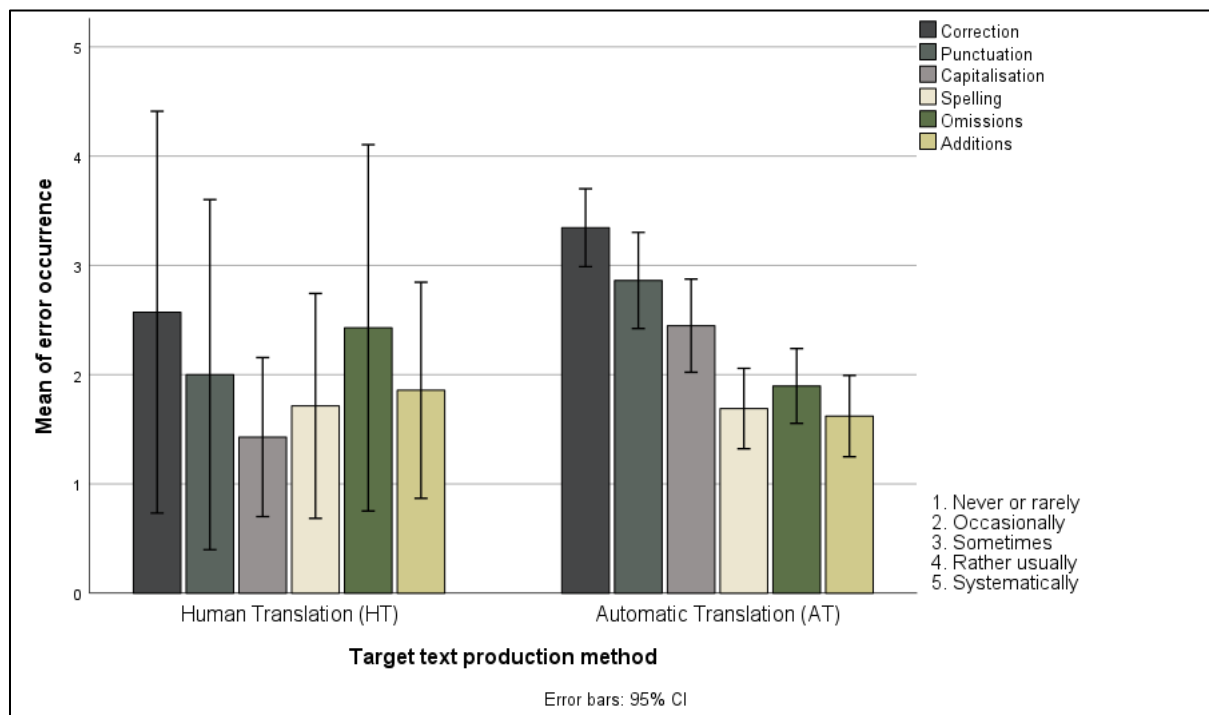
Table 5
Frequency values for linguistic errors

N=29 (AT) N=7 (HT)	Mean μ		Median M		Std Deviation σ		Variance σ^2		Minimum		Maximum	
	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT	AT	HT
Correction	3.34	2.57	3	1	0.936	1.988	0.877	3.952	1	1	5	5
Punctuation	2.86	2.00	3	1	1.156	1.732	1.337	3.000	0	1	5	5
Capitalization	2.45	1.43	3	1	1.121	0.787	1.256	0.619	1	1	5	3
Spelling	1.69	1.71	1	1	0.967	1.113	0.936	1.238	0	1	3	4
Omissions	1.9	2.43	2	2	0.900	1.813	0.81	3.286	1	1	4	5
Additions	1.62	1.86	1	2	0.979	1.069	0.958	1.143	1	1	4	4

A Mann-Whitney U test ($\alpha=0.05$) was performed, revealing that only the capitalization variable could be statistically significant when considering groups AT and HT ($p=.029$). As illustrated in Figure 4, the 0.05 margin of error shows appreciably higher values in the case of the HT because of the standard deviation present in each of the categories:

Figure 4

Linguistic errors identified by the AT and HT groups



As happened in the case of the technical errors, the perception of the HT seems slightly more positive than that held about its AT counterparts. Yet the scores could have been higher if the sample had comprised a greater number of participants; conversely, the scores would have been lower if participant 3.14 (i.e., the outlier) had been discarded from the sample. As suggested by the skewness variation among the HT participants, especially in the case of omissions and additions, it can be established that for some respondents the HT deviated greatly from the original English template in its content and wording. The standard deviation is significantly lower for capitalization and spelling, two categories that were also rated positively among the AT participants.

The spread of the errors identified in the HT and AT templates is not directly conducive to assuming that there is a correlation between the AT and a higher number of errors, or between the HT and a lower number of errors. This is contrary to the statistically significant values of the overall quality variables, which suggest that the text production method has a bearing on the participants' scoring of each variable. Simply stated, the data do not allow us to suggest that the errors caused by the NMT engine are systematically perceived as being worse than those caused by the human subtitler.

4.3. Opinions on subtitle post-editing as a new skill of professional value

After evaluating the templates, the participants were prompted to reflect on their own skills and performance. As previously established in the sample description, most of them (22; 61%) had undergone some sort of post-editing training during their university studies, but only one (3%) had had substantial post-editing experience. The aim was not only to understand how well equipped they felt when undertaking this task, but also to gauge their opinion of post-

editing as a professional practice (and skill) that is becoming more common in professional scenarios.

The respondents were also asked to evaluate their own subtitling, translation, and post-editing skills. On a Likert scale ranging from 1 (“completely disagree”) to 5 (“completely agree”), they could choose to indicate how confident they felt in those three hard skills when completing the task. As can be seen in Table 6, the participants were overly positive about self-evaluating their skills, especially their translation competence, with a high median and minimal deviation:

Table 6
Frequency values for self-evaluation of skills

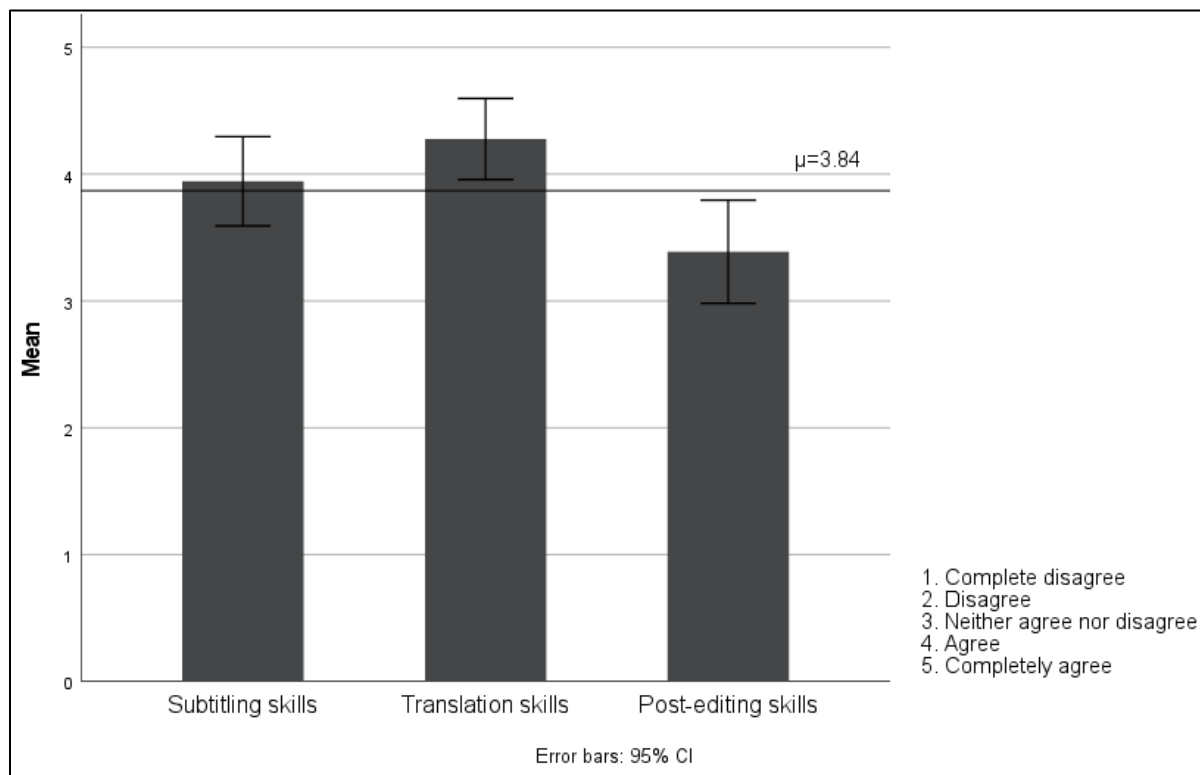
N=36	Mean μ	Median M	Std Deviation σ	Variance σ^2	Minimum	Maximum
1. Self-evaluation of subtitling skills	3.94	4.00	1.040	1.083	0	5
2. Self-evaluation of translation skills	4.28	4.00	0.944	0.892	0	5
3. Self-evaluation of post-editing skills	3.39	4.00	1.202	1.444	0	5

Values are not normally distributed in the case of the first two questions, where a negative skewness of more than -2.140 and -2.757 , respectively, can be observed. These values are high because most responses are indeed around the maximum value ($m=4$), except for a few lower scores that detract from the means. To explain this abnormality, we observe that one respondent selected *N/A* for *translation skills* and two chose *N/A* and *completely disagree* for *subtitling skills*.

Figure 5 shows minimal error margins in each category, confirming that there is general agreement on the positive self-evaluation of their subtitling, translation, and post-editing skills. It is clear, however, that the students feel less confident in post-editing, which is surely influenced by the significantly less training received (or complete lack of it), which was mentioned in the previous section.

Figure 5

Self-evaluation of subtitling, translation, and post-editing skills by all groups



The participants were also asked to rate their agreement or disagreement with some adjectives describing the subtitle post-editing task at hand, irrespective of the group to which they belonged. The low standard deviation values may indicate that the responses of the HT group were not so considerably different as to cause substantial changes. As shown in Table 7, the responses to the first two adjectives were positive (i.e., *manageable* and *interesting*), whereas the responses to the other six attributes were overtly negative. The data obtained from all the questions show a normal distribution, with skewness values between -1.473 and 0.144 :

Table 7

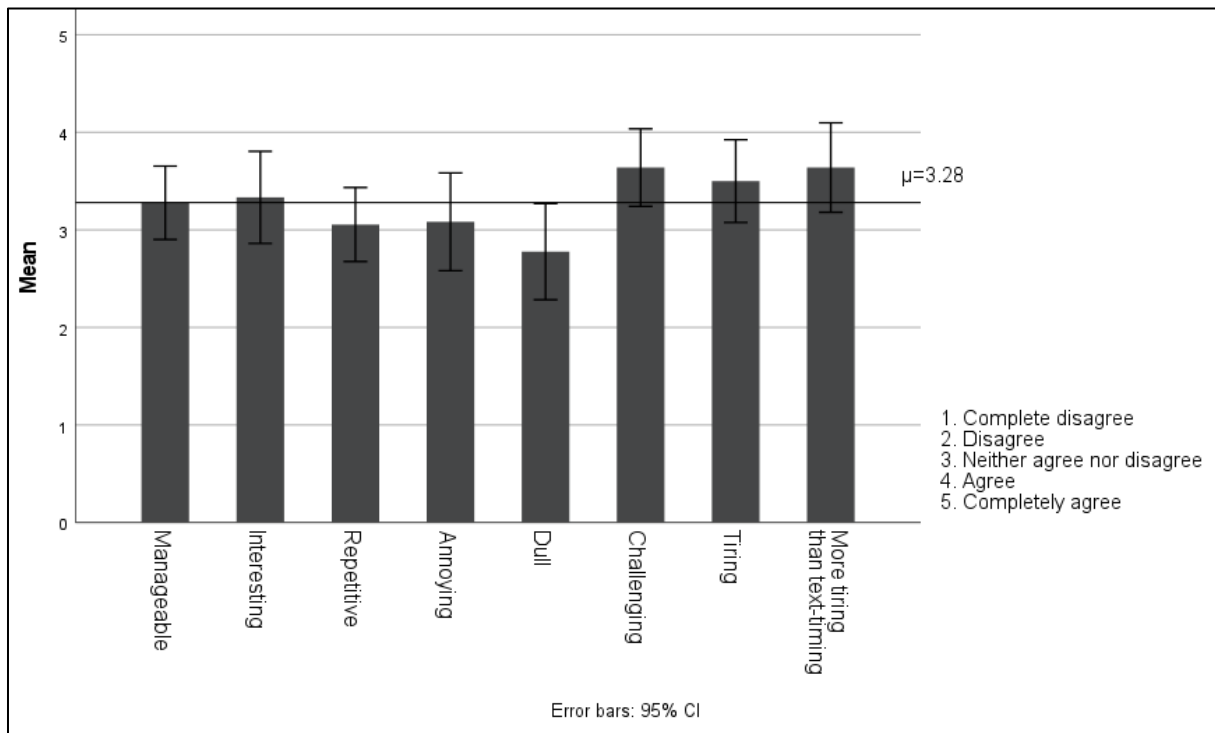
Frequency values for adjectives used to describe subtitle post-editing

Subtitle post-editing is:	Mean μ	Median M	Std Deviation σ	Variance σ^2	Minimum	Maximum
1. Manageable	3.28	4.00	1.111	1.235	0	5
2. Interesting	3.33	4.00	1.394	1.943	0	5
3. Repetitive	3.06	3.00	1.120	1.254	1	5
4. Annoying	3.08	3.00	1.481	2.193	0	5
5. Dull	2.78	3.00	1.456	2.121	0	5
6. Challenging	3.64	4.00	1.175	1.380	0	5
7. Tiring	3.50	3.50	1.254	1.571	0	5
8. More tiring than timing text from scratch	3.64	4.00	1.355	1.837	1	5

The visual illustration of the above results, as seen in Figure 6, shows that the mean values for most questions are relatively similar ($\mu=3.28$), signalling that most of the participants were in agreement with the adjectives used to describe the task:

Figure 6

Description of subtitle post-editing (AT and HT groups) after the experiments



If anything, there is a tendency to consider subtitle post-editing as *challenging*, *tiring* and *more tiring than text-timing*, with the positive adjectives being slightly lower on the Likert scale. The lowest score was for *dull* ($\mu=2.78$), which displays visible disagreement with this and confirms the higher score for *interesting* ($\mu=3.33$). The participants' perception of subtitle post-editing, while visibly unfavourable, is more positive than initially expected when designing the experiment, with the median for positive adjectives being higher than the median for negative ones (4 and 3, respectively) – the only two exceptions being *challenging* and *more tiring than text-timing*. The participants' opinions on the latter adjective might indicate a preference to translate subtitles from scratch rather than using a pre-translated template, but this calls for further examination.

In the last part of the questionnaire, the respondents shared their insights into subtitle post-editing as a professional practice that is gaining prominence in the AVT industry. As displayed in Table 8, the data are distributed normally along the spread and show skewness values of between -1.274 and 0.050 :

Table 8

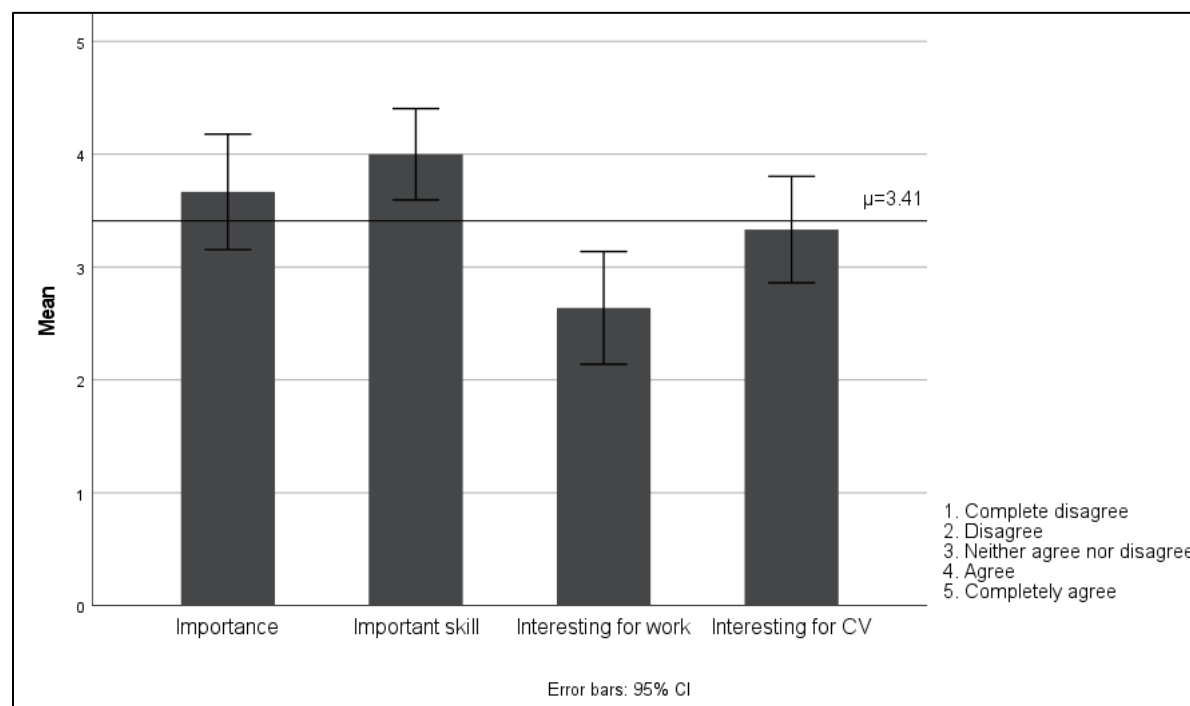
Frequency values for the perception of subtitle post-editing as a professional practice

N=36	Mean μ	Median M	Std Deviation σ	Variance σ^2	Minimum	Maximum
1. Subtitle post-editing is important nowadays	3.67	4.00	1.512	2.286	0	5
2. Subtitle post-editing is a good skill to have	4.00	4.00	1.195	1.429	1	5
3. I would do subtitle post-editing professionally in future	2.64	3.00	1.477	2.180	0	5
4. I would add subtitle post-editing to my CV	3.33	4.00	1.394	1.943	0	5

The median of these questions is roughly similar throughout ($m=4$, *agree*) except for the participants' willingness to do subtitle post-editing professionally in the future, which has a visibly lower median ($m=3$) and mean ($\mu=2.64$); this suggests that would-be subtitlers are not too eager to engage in this activity professionally. Figure 7 is a visual representation of the above information, implying that the respondents are not keen to undertake subtitle post-editing professionally even if they consider it important in today's industry landscape:

Figure 7

Perception of subtitle post-editing in today's professional landscape



This section has discussed the results obtained from the statistical analysis conducted on a sample of students ($N=36$) comprising two distinct groups ($AT=29$ and $HT=7$). The results have demonstrated the presence of statistical significance on certain variables when comparing both groups, especially the overall perception of quality, which is significantly higher for HT than AT subtitles across the board (i.e., linguistic, translation, and technical). Having said that, differences were less sharp regarding the identification of specific errors, whether about their

nature, severity, or frequency. Yet, after a sig. two-tailed correlation test, no significant differences have been observed between the groups regarding their perception of post-editing – which in both cases is rather negative. This finding highlights a paradoxical situation in which would-be subtitlers are aware of market needs and try to acquire new skills and yet they are ultimately reluctant to engage in this professional activity.

There are limitations to this study, starting with the relatively small number of participants. Other educational settings, language combinations, and NMT engines would help to offer a more holistic view of subtitle post-editing. Moreover, it can be argued that prejudice about post-editing as a professional practice is directly proportional to a lack of training, so surveying students with more experience in the topic might yield different results.

5. Conclusions

The post-editing of subtitles is a novel practice on which little research has been conducted to date, particularly regarding the perceptions and appreciation of would-be AVT professionals. When approaching the design and rollout of the experiments, our initial hypothesis was that the participants would have a negative perception of post-editing, most likely exacerbated by their exposure to a machine-generated subtitle template during one of the experiments. In this respect, our main hypothesis has been confirmed since the participants in the AT groups found the quality of the AT templates poor and their opinion of subtitle post-editing was accordingly negative. The results also show that the HT group conferred significantly higher scores on the global quality of the subtitle template regarding translation, technical, and linguistic dimensions and, despite this positive appreciation, their overall opinion on post-editing was equally unfavourable and similar to that of the AT groups. A negative predisposition and the fact that post-editing evaluation taxonomies are prone to capitalize on errors (O'Brien, 2011) are factors that might have affected some of the responses.

The participants were fully aware that post-editing is becoming used with increasing frequency in the language industry (Robert et al., 2023) and acknowledged that it is an interesting and much-needed skill to acquire in today's AVT market. However, they also expressed their reluctance to engage in this activity themselves. This paradox suggests a mismatch between what would-be subtitlers think they must learn and what they would actually like to do instead. These findings echo those obtained by Moorkens et al. (2018), who found it paradoxical that while most post-editors considered NMT to be efficient, they were nevertheless reluctant to embrace post-editing and preferred to translate from scratch for the sake of creativity. As this study's participants did not know whether they were dealing with post-editing (AT groups) or revision (HT group), the inference can be made that they are unenthusiastic not only about engaging in post-editing, but also perhaps about engaging in editing altogether (including revision).

The participants' theoretical knowledge and practical know-how of post-editing were considered to be somewhat limited at the time of the experiments, and their opinions about this novel practice and that of the more traditional revision are well aligned. It can be argued that the jury is still out on to how the future generation of AVT professionals perceives post-editing. The limitations of this study call for further research that can help to establish correlations between the level of training and the perception of post-editing.

Bolaños García-Escribano, A., & Díaz-Cintas, J. (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137

Few would deny the pedagogical challenge that novel practices, such as subtitle post-editing, pose to translator educators. Yet technology-led professional practices are here to stay for the foreseeable future and perception studies are therefore crucial to informing training practices.

Funding

This research was funded by a “Margarita Salas” postdoctoral research fellowship from Universitat Jaume I (ref. no. MGS/2022/03), financed by the European Union – Next Generation EU.

References

- Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 297–317). John Benjamins. <https://doi.org/10.1075/btl.35.19all>
- Athanasiadi, R. (2017). The potential of machine translation and other language assistive tools in subtitling: A new era? In M. Deckert (Ed.), *Audiovisual translation: Research and use* (pp. 29–49). Peter Lang.
- ATRAE. (2021). Comunicado sobre la posesición. Asociación de Traducción y Adaptación Audiovisual de España. <https://atrae.org/comunicado-sobre-la-posedicion>
- Bogucki, Ł., & Deckert, M. (Eds.) (2020). *The Palgrave handbook of audiovisual translation and media accessibility*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-42105-2>
- Bolaños García-Escribano, A., & Declercq, C. (2023). Editing in audiovisual translation (subtitling). In S.-W. Chan (Ed.), *The Routledge encyclopedia of translation technology* (pp. 565–581). Routledge. <https://doi.org/10.4324/9781003168348-36>
- Bolaños García-Escribano, A., & Díaz-Cintas, J. (2020). The cloud turn in audiovisual translation. In Ł. Bogucki & M. Deckert (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 519–544). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-42105-2_26
- Bolaños García-Escribano, A., Díaz-Cintas, J., & Massidda, S. (2021). Subtitlers on the cloud: The use of professional web-based systems in subtitling practice and training. *Tradumàtica*, 19, 1–21. <https://doi.org/10.5565/rev/tradumatica.276>
- British Standards Institute (BSI). (2015). *Translation services – Post-editing of machine translation – Requirements*.
- Burchardt, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2016). Machine translation quality in an audiovisual context. *Target*, 28(2), 206–221. <https://doi.org/10.1075/target.28.2.03bur>
- Bywood, L., Georgakopoulou, P., & Etchegoyhen, T. (2017). Embracing the threat: Machine translation as a solution for subtitling. *Perspectives*, 25(3), 492–508. <https://doi.org/10.1080/0907676X.2017.1291695>
- Columbus, C. (Director). (1993). *Mrs Doubtfire* [Film]. Twentieth Century Fox; Blue Wolf Productions.
- de Sousa, S., Aziz, W., & Specia, L. (2011). Assessing the post-editing effort for automatic and semi-automatic translations of DVD subtitles. *Proceedings of Recent Advances in Natural Language Processing* (pp. 97–103). Association for Computational Linguistics. <https://aclanthology.org/R11-1014>
- Declercq, C. (2023). Editing in translation technology. In S.-W. Chan (Ed.), *The Routledge encyclopedia of translation technology* (pp. 551–564). Routledge. <https://doi.org/10.4324/9781003168348-35>

- Bolaños García-Escribano, A., & Díaz-Cintas, J. (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137
- Delabastita, D. (1989). Translation and mass-communication: Film and TV. Translation as evidence of cultural dynamics. *Babel*, 35(4), 193–218. <https://doi.org/10.1075/babel.35.4.02del>
- Díaz-Cintas, J. (2023). Technological strides in subtitling. In S.-W. Chan (Ed.), *The Routledge encyclopedia of translation technology* (pp. 720–730). Routledge. <https://doi.org/10.4324/9781003168348-46>
- Díaz-Cintas, J., & Massidda, S. (2019). Technological advances in audiovisual translation. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 255–270). Routledge. <https://doi.org/10.4324/9781315311258-15>
- Díaz-Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices*. Routledge. <https://doi.org/10.4324/9781315674278>
- do Carmo, F., Shterionov, D., Moorkens, J., Wagner, J., Hossari, M., Paquin, E., Schmidtke, D., Groves, D., & Way, A. (2021). A review of the state-of-the-art in automatic post-editing. *Machine Translation*, 35(2), 101–143. <https://doi.org/10.1007/s10590-020-09252-y>
- Easton, J. (2021, April 26). OTT revenues to top US\$200 billion by 2026. *Digital TV Europe*. www.digitaltveurope.com/2021/04/26/ott-revenues-to-top-us200-billion-by-2026/#close-modal
- Federico, M., Cattelan, A., & Trombetti, M. (2012). Measuring user productivity in machine translation enhanced computer assisted translation. *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas: Research papers* (pp. 1–10). <https://aclanthology.org/2012.amta-papers.22>
- Gambier, Y., & Gottlieb, H. (Eds.) (2001). *(Multi) media translation: Concepts, practices and research*. John Benjamins. <https://doi.org/10.1075/btl.34>
- Gaspari, F., Toral, A., Kumar Naskar, S., Groves, D., & Way, A. (2014). Perception vs reality: Measuring machine translation post-editing productivity. In S. O'Brien, M. Simard & L. Specia (Eds.), *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas* (pp. 60–72). Association for Machine Translation in the Americas. <https://aclanthology.org/2014.amta-wptp.5>
- Georgakopoulou, P., & Bywood, L. (2014). MT in subtitling and the rising profile of the post-editor. *Multilingual*, 25(1), 24–28. <https://multilingual.com/articles/mt-in-subtitling-and-the-rising-profile-of-the-post-editor>
- González Pastor, D. (2021). Introducing machine translation in the translation classroom: A survey on students' attitudes and perceptions. *Tradumàtica*, 19, 47–65. <https://doi.org/10.5565/rev/tradumatica.273>
- Guerberof Arenas, A. (2008). *Productivity and quality in the post-editing of outputs from translation memories and machine translation* [Doctoral dissertation, Universitat Rovira i Virgili]. Tesis Doctorals en Xarxa. <http://hdl.handle.net/10803/90247>
- Gupta, P. Sharma, M., Pitale, K., & Kumar, K. (2019). Problems with automating translation of movie/TV show subtitles. *ArXiv*. <https://doi.org/10.48550/arXiv.1909.05362>
- Hu, K., & Cadwell, P. (2016). A comparative study of post-editing guidelines. *Baltic J. Modern Computing*, 4(2), 346–353. <https://aclanthology.org/W16-3420>
- Koponen, M. (2012). Comparing human perceptions of post-editing effort with post-editing operations. In C. Callison Burch, P. Koehn, C. Monz, M. Post, R. Soricut & L. Specia (Eds.), *Proceedings of the Seventh Workshop on Statistical Machine Translation* (pp. 181–190). Association for Computational Linguistics. <https://aclanthology.org/W12-3123>
- Koponen, M. (2016). *Machine translation post-editing and effort: Empirical studies on the post-editing process* [Doctoral dissertation, University of Helsinki]. University of Helsinki open repository. <http://hdl.handle.net/10138/160256>
- Koponen, M., Sulubacak, U., Vitikainen, K., & Tiedemann, J. (2020a). MT for subtitling: User evaluation of post-editing productivity. *Proceedings of the 22nd Annual Conference of the European*

- Bolaños García-Escribano, A., & Díaz-Cintas, J. (2023) Integrating post-editing into the subtitling classroom: What do subtitlers-to-be think? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 22, 115–137
- Association for Machine Translation (pp. 115–124). European Association for Machine Translation. <https://aclanthology.org/2020.eamt-1.13>
- Koponen, M., Sulubacak, U., Vitikainen, K., & Tiedemann, J. (2020b). MT for subtitling: Investigating professional translators' user experience and feedback. *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas* (pp. 79–92). Association for Machine Translation in the Americas. <https://aclanthology.org/2020.amta-pemdt.6>
- Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent State University Press.
- Läubli, S. & Orrego-Carmona, D. (2017). When Google Translate is better than some human colleagues, those people are no longer colleagues. *Translating and the Computer*, 39, 59–69. <https://doi.org/10.5167/uzh-147260>
- Matusov, E., Wilken, P., & Georgakopoulou, P. (2019). Customizing neural machine translation for subtitling. *Proceedings of the Fourth Conference on Machine Translation* (pp. 82–93). Association for Computational Linguistics. <http://doi.org/10.18653/v1/W19-5209>
- Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2), 240–262. <https://doi.org/10.1075/ts.18014.moo>
- Mossop, B. (2020). *Revising and editing for translators*. Routledge. <https://doi.org/10.4324/9781315158990>
- O'Brien, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation*, 19(1), 37–58. <https://doi.org/10.1007/s10590-005-2467-1>
- O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation*, 25(3), 197–215. <https://doi.org/10.1007/s10590-011-9096-7>
- Pérez-González, L. (Ed.) (2018). *The Routledge handbook of audiovisual translation*. Routledge. <https://doi.org/10.4324/9781315717166>
- Pérez Macías, L. (2020). ¿Qué piensan los traductores sobre la posesición?: Un estudio basado en el método mixto sobre los temores, las preocupaciones y las preferencias en la posesición de traducción automática. *Tradumàtica*, 18, 11–32. <https://doi.org/10.5565/rev/tradumatica.227>
- Plitt, M., & Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics*, 93, 7–16.
- Robert, I., Ureel, J. J. J., & Schrijver, I. (2023). Translation, translation revision and post-editing competence models: Where are we now? In G. Massey, E. Huertas-Barros & D. Katan (Eds.), *The human translator in the 2020s* (pp. 44–59). Routledge. <https://doi.org/10.4324/9781003223344-4>
- Romero-Fresco, P., & Chaume, F. (2022). Creativity in audiovisual translation and media accessibility. *The Journal of Audiovisual Translation*, 38, 75–101.
- Rossi, C., & Chevrot, J.-P. (2019). Uses and perceptions of machine translation at the European Commission. *The Journal of Specialised Translation*, 31, 177–200.
- Specia, L., & Farzindar, A. (2010). Estimating machine translation post-editing effort with HTER. In V. Zhechev (Ed.), *Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry* (pp. 33–41). Association for Machine Translation in the Americas. <https://aclanthology.org/2010.jec-1.5>