



Expertise in interlingual subtitling: applying the FAR model to study the quality of subtitles created by professional and trainee subtitlers

Ewa Ludera
University of Warsaw, Poland
ewa.ludera@wp.pl

Agnieszka Szarkowska
University of Warsaw, Poland
a.szarkowska@uw.edu.pl

David Orrego-Carmona
University of Warwick, United Kingdom/University of the Free State, South Africa
david.orrego-carmona@warwick.ac.uk

DOI: 10.12807/ti.116201.2024.a04

Abstract: Subtitling quality is a rather slippery notion and its assessment in interlingual subtitling continues to present a challenge to subtitling trainers, broadcasters, language service providers and other stakeholders. Using unexplored data from a subtitling process research study by Orrego-Carmona et al. (2018), we examined the quality of interlingual English-to-Polish subtitles created by professionals and novices. First, we implemented Pedersen's (2017) FAR model to assess the quality of subtitles created by the participants, predicting that professionals would achieve higher quality scores than novices. Then, we followed up the FAR model examination with a quantitative analysis of a set of quality parameters related to text condensation, which is considered a key skill in interlingual subtitling. Despite our hypotheses, professional subtitlers in our study did not gain higher scores in the FAR model analysis; they also made similar types of errors as novices. However, their expertise was demonstrated in better condensation skills. We have also discovered an interesting relationship between subtitlers' age and their condensation skills. Our study may contribute to a fuller understanding of expertise in interlingual subtitling and provide subtitling trainers with clues about areas most problematic for students.

Keywords: Subtitling; quality; expertise; eye-tracking; FAR model

1. Introduction

Media localisation companies and subtitling trainers across the world are struggling to achieve high-quality subtitling to ensure the best possible viewer experience for mass audiences watching an increasing number of hours of audiovisual content every day. At the same time, more and more viewers are expressing their dissatisfaction with the quality of many audiovisual translations, with most (in)famous cases of subtitling blunders making the headlines (Namkung, 2021). But how do we identify a good-quality translation? Can subtitling quality be measured and parametrised? What makes a good

subtitler and which subtitling skills are the most problematic for novices? And finally, what distinguishes experienced subtitling professionals from trainees?

With these questions in mind, in this paper, we take a closer look at the quality of interlingual subtitles created by professional and trainee subtitlers using previously unexplored data collected by Orrego-Carmona et al. (2018) in their translation process research study. After providing an overview of approaches to quality assessment in subtitling, we use Pedersen's (2017) FAR model to assess the quality of subtitles created by novice and professional subtitlers. The model focuses specifically on interlingual subtitling, making it more suitable than other quality assessment models in translation (Castilho et al. 2018). Then, we follow up the FAR model analysis with a quantitative analysis of a set of subtitling quality parameters related to condensation. We hope that the results of our analyses will contribute to a fuller understanding of expertise in interlingual subtitling as well as provide subtitling trainers with helpful information about areas which may prove problematic for students.

2. Expertise in translation

As aptly noted by Hurtado Albir (2017), translation is a “communicative activity that involves decision-taking, problem-solving and, like other similar activities, requires expert knowledge” (p. 3). The most popular models of translation competence, such as PACTE or EMT, conceptualise competence as a multifaceted notion consisting of several sub-competences or components (EMT, 2017; Hurtado Albir, 2017). Together, the combination of these components makes up translation expertise, which ultimately allows us to distinguish translators from non-translators.

Expertise can be defined as “the characteristics, skills and knowledge that distinguish experts from novices and less experienced people” (Ericsson et al. 2006, p. 3). Ericsson (2018, p. 696) notes that the word “expert” has the same origin as “experience”, which may suggest a positive correlation between experience and level of expertise. Tiselius and Hild (2017, p. 425) point out that, while competence and expertise are commonly confused in translation and interpreting studies, expertise is the “supreme expression” and mastery of a pre-existing competence. Expertise is built over years through focused work and deliberate practice. Given their extensive previous experience, experts are expected to achieve higher levels of performance than trainees who are only starting to use their newly acquired competence. In this study we set out to find whether professional subtitlers outperform trainees in an interlingual subtitling task.

2.1 Previous research on translation expertise

Expertise can be explored from various angles through a range of different research approaches, such as concurrent, retrospective or longitudinal studies (Baker et al., 2018). In translation and interpreting studies, the “majority of the studies on competence/expertise were designed as expert–novice comparisons” (Tiselius and Hild, 2017, p. 434) and have adopted a retrospective or longitudinal perspective. In those studies, translations produced by professionals usually served as a reference for evaluating those done by students or by bilinguals with no translation experience (see, for instance, Christoffels et al., 2006). Some of the studies also attempted to pinpoint the characteristics of translation expertise. For instance, a translation process research study conducted by Whyatt (2018), in which professional translators, trainee translators and language students carried out a translation and a paraphrasing task, found that professionals used online resources less than the other two

groups, did not make as many long pauses, typed faster on a keyboard and produced higher-quality translations. However, the difference between target text quality produced by professional and trainee translators was not statistically significant, which indicates that the number of years of experience is not the only important factor influencing translation quality. In this study, quality was assessed by two experienced translator trainers using two scales: a holistic one and an error detection-based one (Whyatt, 2018).

In a longitudinal study conducted by Göpferich (2009), a group of twelve translation students were asked to translate ten texts from English into German – their first language – at the beginning of each semester of their three-year studies and at the end of their last semester. The comparison of their translations with translations done by ten professionals and of the translation process showed that professionals made more overall successful decisions and more low-effort decisions than students and that their low-effort decisions had a considerably higher success rate. By contrast, the professionals' success rate in high-effort decisions was only approximately 2% higher than that achieved by students (Göpferich, 2013).

In another longitudinal study conducted by Chmiel (2021), a group of interpreting trainees was tested at the beginning and the end of their training, and their end-of-training results were compared to results obtained by professional interpreters. The study found that the accuracy rate, i.e., the percentage of correct renditions, achieved by trainees was approximately 87% before and 93% after training, which could indicate that accuracy increases with training. The difference in results obtained by advanced trainees and professionals appears to be rather small, as does the translation speed.

2.2 Previous research on expertise in subtitling employing process research

In contrast to translation and interpreting studies, experimental or descriptive studies evaluating expertise in subtitling only started to appear in recent years and the number of studies remains limited. Based on extensive participant observation, questionnaires, screen recording and interviews, Beuchert (2017) reconstructed the process of subtitling in the Danish context through the study of five professional subtitlers. Beuchert provided a detailed account of the internal, external and intersectional elements that interact in the creation of subtitles and a comprehensive report of the competences, tasks and skills required to complete subtitling tasks.

In another study, Tardel (2020) compared professionals and students to assess the integration of automatic speech recognition (ASR) into semi-automated subtitling processes. The results do not suggest a benefit from integrating ASR, which Tardel attributes to the quality of the automatic translation. Regarding production effort in the translation task, there were no significant differences in terms of time invested in completing the task, but professionals worked more efficiently when it came to technical effort.

A translation process research study using eye tracking conducted by Orrego-Carmona et al. (2018) examined the process of creating interlingual subtitles by professional and trainee subtitlers. Focussing on the *process* of creating subtitles, Orrego-Carmona et al. (2018) did not include a qualitative analysis of the translation *product*: the output produced by study participants. This paper uses data collected by Orrego-Carmona et al. (2018) and examines various aspects of the quality of subtitles created by professional subtitlers and subtitling trainees using the FAR model.

3. Subtitling quality

Quality is a multi-faceted notion that can be understood in different ways depending on the context and the stakeholders involved (Castilho et al., 2018). The evaluation of translations has been essential in the consolidation of translation studies, and the systematization of quality assessment has been essential to the consolidation of the localisation industry. With the expansion of the industry, quality has moved from theory-driven models to industry-operationalisable models. Interlingual subtitling quality has not received much attention. According to Künzli, only 11 entries in the Translation Studies Bibliography between 2009 and 2019 deal with this topic (2021, p. 326). When it comes to the perception of quality among different stakeholders, a questionnaire study by Robert and Remael (2016) revealed that subtitlers tended to focus on multiple aspects of quality, whereas language service providers (LSPs) paid more attention to technical aspects. Surveying professional subtitlers and viewers, Szarkowska et al. (2021) found that while subtitlers tended to emphasise the importance of synchronisation, condensation and idiomaticity of subtitles, viewers often focussed on the discrepancies between the dialogue and the subtitles.

The quality in interlingual subtitling can be approached from two perspectives: *ensuring* and *assessing* quality. Ensuring quality takes place before the subtitling process and consists in providing the subtitler with the requirements on how the task should be performed, typically through subtitling guidelines. Assessing quality, on the other hand, occurs after the completion of the subtitling process and consists in evaluating the finished product, typically at the quality-checking stage.

3.1. Ensuring subtitling quality

Probably the most common way of ensuring quality by content providers and LSPs is subtitling guidelines (see, for instance, those by Netflix (2021a, 2021b, 2021c, 2021d), Channel 4 (n.d.), ARTE (2019) or TED (2017)). These guidelines are intended to standardise the production processes in terms of technical parameters and to ensure consistency – a proxy for the high quality of the product. By providing various subtitling metrics, such as reading speed, text segmentation or line length, style guides can be used to gauge quality as well as prevent errors from occurring. This, in turn, is supposed to ensure quality by reducing the time and effort needed for revision at the quality control stage.

With globalisation, some subtitling recommendations are applied universally across languages and countries, however, recommendations are also influenced by national traditions (Díaz-Cintas & Remael, 2014; ESIST, 2021). With the expansion and professionalisation of AVT, national guidelines have been put in place to reflect those traditions. Multiple sets of such guidelines were gathered by ESIST (2021) on its website. Pedersen (2018) notes that national guidelines have evolved as technology and viewing habits changed. In countries with the predominantly subtitling tradition, such as Sweden, Denmark or Greece, company guidelines were developed based mainly on subtitling practices in public-service broadcasters and later were adopted as national guidelines. Developed in isolation, they were generally not influenced by other national guidelines. In consequence, noticeable differences between various national guidelines have consolidated and hence subtitles produced in different countries may be largely different.

3.2. Assessing subtitling quality: the FAR model

Despite the difficulty in defining quality, operationalising this concept allows for its evaluation. After all, as Pedersen (2017, p. 210-211) posits, “many people

have to judge translation quality on a daily basis: revisers, editors, evaluators, teachers, not to mention the subtitlers themselves, and of course: the viewers”. Media localisation companies also use their own QC criteria on a daily basis to eliminate errors. For instance, Netflix (n.d.-a, n.d.-b, n.d.-c) quality criteria cover most common subtitling errors, including technical ones, e.g., “Sync – Global Offset”, which concerns all subtitles being out of sync by the same amount of time, linguistic ones, e.g. “Translation – SGP” (spelling, grammar, punctuation), and translational ones, e.g. “Missing Content – Conversation Event”, which concerns not subtitling dialogues. Interestingly, Netflix uses a method of calculating the error rate which consists in dividing the number of subtitles with errors by the total number of subtitles. This error rate is used to evaluate subtitlers.

When it comes to subtitling quality assessment, a few models have been developed. Two of them concern live subtitling: the NER model (Romero-Fresco & Pérez, 2015) and the NTR model (Romero-Fresco & Pöchhacker, 2017). The NER model served as the foundation for the FAR model developed by Pedersen (2017), which we adopted in our study and which we describe below in more detail.

A recent proposal to assess subtitle quality was put forward by Künzli (2021). The CIA model to evaluate subtitle quality integrates three dimensions (Correspondence, Intelligibility and Authenticity) and was developed based on a survey of 59 professional subtitlers. The model proposes to assess these three dimensions and hypothesises that, when these assumptions are met, “the reception of a subtitled audiovisual production will lead to a flow experience” (Künzli, 2021, p. 336). However, this model is yet to be tested empirically and for the comparability and assessment purposes of our analysis, the FAR model remains the most suitable alternative for our study.

The FAR model is the only publicly available assessment model for interlingual subtitles prepared before airing. The basic units of assessment are single subtitles, which are evaluated with regard to three areas: functional equivalence, acceptability and readability (Pedersen, 2017). The model represents a viewer-oriented and a product-oriented approach and focuses on error analysis. It introduces a three-tier penalty point system: minor (0.25 points), standard (0.5 points) and serious (1 point). Error severity is chosen based on a concept which Pedersen (2017, p. 215) refers to as the “contract of illusion”: the viewer should be able to forget that they are reading a translation in the form of subtitles, which should be considered as the dialogue itself. Minor errors usually do not break the contract of illusion, as they are only noticeable to particularly attentive viewers; standard errors are generally visible to the viewer, they break the contract and disturb the watching experience; serious errors are so obvious that they make it difficult to understand the subtitle in which they occur and the following ones.

Errors in the functional equivalence area can be either semantic or stylistic. Semantic errors can cause the most issues with correctly understanding the plot; thus, they are penalised more severely than other types of errors: 0.5 points for minor, 1 point for standard and 2 points for serious errors. Minor semantic errors usually involve choosing the wrong word, but these errors do not affect the understanding of the plot. Standard errors of this kind result in a loss of meaning in the subtitle in which they occur, e.g., verbatim translations or some omissions. Serious errors impede the comprehension of the following subtitles. Stylistic errors include, among others, choosing the wrong register or using anachronisms.

Acceptability concerns grammar, spelling and idiomaticity errors. Minor grammar errors do not affect the understanding of the subtitle, serious ones make it challenging to make sense of the subtitle, whereas standard errors fall

somewhere in between the two. Spelling errors are minor if they do not change the meaning of the word, standard if they change the meaning of the word and serious if they make the word impossible to decode. Idiomaticity errors make the translation sound unnatural to native speakers.

Readability concerns segmentation and spotting, punctuation and graphics, reading speed and line length. Spotting errors occur when subtitles are not synchronised with the dialogue. They are classified as minor when the subtitle is out of sync by less than one second, as standard when it is off by at least one second, but no more than one utterance, and serious when it is off by more than one utterance. Minor segmentation errors occur if the line break is incorrectly placed within one subtitle, whereas standard errors of this kind occur if the segmentation between consecutive subtitles is incorrect; segmentation errors are typically not serious. Errors in the punctuation and graphics category may concern, for instance, the incorrect use of dashes or italics: in these cases, they are classified as standard. Pedersen (2017) states that reading speed errors depend on the guidelines followed. However, he also proposes his own classification: minor errors concern reading speeds above 15 characters per second (cps), whereas standard errors occur when a subtitle has a reading speed of 20 cps. Pedersen claims that this is because with a reading speed of 15 cps, viewers already spend most of the time looking at subtitles rather than the image, whereas with a reading speed of 20 cps they would likely not be able to focus on the image at all.

Pedersen (2017) suggests a method of calculating scores for the three areas separately: the sum of penalty points should be divided by the number of subtitles. Total scores can be calculated by adding all the penalty points for the three areas and then dividing the sum by the number of subtitles. As Pedersen (2017) notes, the FAR model can easily be expanded and modified according to one's own needs. What is more, it may constitute a useful tool for providing feedback and for teachers.

4. Overview of the current study

Studying the performance of trainees and professionals can show us whether there are any discernible differences between them in terms of the quality of subtitles they create. This, in turn, may contribute to improving subtitling training, as it may indicate areas with which trainees struggle most and which may require additional attention in class. With this goal in mind, we decided to apply the FAR model to study the quality of interlingual subtitles created by professional subtitlers and subtitling trainees, using the unexplored data from the study reported in Orrego-Carmona et al. (2018). Their study focused on the translation process aspects and did not include any analysis of the qualitative translational aspects of the subtitles.

The original study used eye tracking and key logging to explore three types of effort: temporal (operationalised as task completion time), cognitive (time spent on Internet research and two types of eye-tracking data: mean fixation duration and dwell time) and production (mouse clicks, keylogging and text reduction). The participants were asked to time and translate subtitles from English into Polish using the professional subtitling software they normally work with, either EZTitles or EdList. It was found that – as expected – professional subtitlers completed the task faster than trainees. They also spent a higher percentage of the time doing online research.

Interestingly, the study by Orrego-Carmona et al. (2018) revealed that professional subtitlers do not necessarily constitute a very homogeneous group: they work with different subtitling tools (in this case EdList vs EZTitles), they

have different educational backgrounds, subtitling training and experience in different subtitling scenarios (national TV broadcasters vs. streaming providers) and work with different style guides (with lower or higher reading speeds). All this may have a bearing on how they perceive the subtitling profession and what constitutes good quality in subtitling. The above differences also make it more difficult to disentangle the impact of subtitling tools, such as EZTitles vs. EdList, from individual differences, such as previous experience or age. For this reason, in our paper, we conduct the analyses using three rather than two groups: older professionals using EdList, younger professionals using EZTitles and trainees, who also used EZTitles. Unlike the original study, which focussed on the subtitling *process*, this paper analyses the quality of the *product*, i.e., subtitles created by the three groups of participants, using the FAR model.

5. Method

5.1. Participants

The present study analyses data from nine professional subtitlers and five trainees (TR). The nine professionals were recruited through the mailing list moderated by the Polish Association of Audiovisual Translators (STAW). Out of the nine professionals, six used EZTitles (PROEZT; mean age = 29.33, $SD = 6.9$) and three used EdList (PROED; mean age = 48.00, $SD = 7.0$). All five trainees used EZTitles (TR, mean age = 22.60 years, $SD = 0.55$). All trainees were enrolled in the MA translation programme at the Institute of Applied Linguistics at the University of Warsaw. They completed an introductory course on interlingual subtitling consisting of 30 contact hours spanning over four months and were trained using the EZTitles subtitling software. All participants signed an informed consent. They did not receive any remuneration for their time.

Almost all trainee subtitlers stated that they had no professional experience. All three professionals using EdList had more than six years of professional experience. Two professionals using EZTitles also declared having more than six years and one 1-3 years of professional experience.

5.2. Procedure and materials

Participants were invited to the lab where they were tested individually on a lab computer connected to the SMI RED 250 mobile eye tracker, which tracked their eye movements and recorded their keystrokes and mouse clicks. The participants could import the profiles with settings they normally used, allowing them to use their own keyboard shortcuts in the lab setting.

Before the experiment, the participants received instructions with the style guide to follow: the maximum number of lines (two), the maximum number of characters in a line (37) and the maximum reading speed (15 cps). The participants were asked to translate the dialogues in an 85-second clip from the TV series *The Newsroom* (dir. Aaron Sorkin, 2012) and to time the subtitles. They were provided with an English transcription of the dialogues. There was no time limit for completing the task and the participants were allowed to use the Internet. Given its fast-paced nature and dialogue abundance, the clip was a fitting choice as it lent itself well to condensing text, which is essential for the task. Additionally, it reflected the typical audiovisual content – an HBO US drama – often translated from English for the Polish audience.

Although short, the clip from *The Newsroom*, contained several subtitling problems: culture-bound items, a high density of fast dialogues as well as two jokes. The first one is a response to the expletive “Oh, blow me” uttered by an annoyed journalist. The addressee in his response references the vulgar

connotations of this phrase, saying: “I want you to not use that language in front of women and to forever not suggest that image to me”. The second joke in this clip concerns the age of one of the characters. In response to “Jim Harper, my senior producer”, the speaker comments, “Senior? Is he old enough to drive at night?”. When it comes to culture-bound items, multiple categories of this subtitling problem can be found in the clip from *The Newsroom*: units of measurement, both referring to distance (feet) and volume (gallons), the name of the comic strip *Little Orphan Annie*, and two toponyms: the Gulf of Mexico and the Grand Canyon as well as a chrematonym: Deepwater Horizon. The third category of subtitling problems is the high density of fast-paced dialogues. The speech rate was approximately 250 words per minute, which means that the subtitlers had to considerably shorten the dialogues not to exceed the space and time limits, without leaving out crucial information.

5.3. Analyses

We conducted two sets of analyses: first, we used the FAR model to compare the quality of subtitles created by our participants (see 5.3.1.), and second, we examined some key subtitling parameters related to text condensation in subtitling between the groups (5.3.2).

5.3.1. FAR model

Using Pedersen’s (2017) FAR model, we calculated the score for each participant individually. Each subtitle was assessed separately according to the criteria in the FAR model. Then, the score for each area of the model was calculated in line with the recommendations provided by Pedersen (2017), as this method shows how well each subtitler met the requirements of the three separate areas and it allows for the comparison of these areas between subtitlers. The calculation consisted in adding all the penalty points, dividing the sum by the number of subtitles and multiplying it by 100%.

The total score was calculated differently to what is recommended by Pedersen (2017), who suggests that the penalty points for all three areas should be added and the sum should be then divided by the number of subtitles. In our analysis, the total score was calculated in a way which represents the three scores together in a more straightforward way: the scores for the three areas were added and the sum was divided by three.

For this study, a minor modification to the FAR model related to the reading speed criterion was implemented. Pedersen (2017) suggests that reading speed above 15 cps should be considered a minor error, whereas that of 20 cps – a standard error. He does not specify if serious reading speed errors are possible. Because of technical issues with subtitling software, we allowed a 10% tolerance margin: while the maximum reading speed was 15cps, we considered errors as minor if they exceeded 16.5 cps. The threshold for standard errors was the same as suggested by Pedersen (2017), i.e., above 20 cps. If a subtitle had a reading speed of more than 25 cps, it was penalised as a serious error.

Due to technical issues with corrupted subtitle files, we could not assess spotting. In the FAR model, spotting relates only to synchronisation. Because synchronisation errors were unlikely to have occurred in the clip from *The Newsroom* considering the high density of the dialogues, the spotting criterion was not included in the present analysis. Table 1 presents the modified FAR model used in the present analysis.

Table 1: The penalty point system of the FAR Model with adaptations made for the purpose of the present analysis

		Minor error	Standard error	Serious error
Functional equivalence	Semantic errors	choice of the wrong word, without the plot being affected	loss of meaning in one subtitle	complete loss of meaning, hampers the understanding of the following subtitles
		0.5	1	2
	Stylistic errors	not specified ¹	not specified	not specified
		0.25	0.5	1
Acceptability	Grammar errors	ignored by most viewers	falls between minor and serious errors	hampers the understanding and the decoding of the subtitle
		0.25	0.5	1
	Spelling errors	any spelling error	changes the meaning of the word	makes the word unintelligible
		0.25	0.5	1
	Idiomatycity errors	not specified	not specified	not specified
		0.25	0.5	1
Readability	Segmentation errors	within one subtitle	between subtitles	⁻²
		0.25	0.5	-
	Punctuation and graphics errors	not specified	incorrect use of dashes and "irrealis" italics	not specified
		0.25	0.5	1
	Line length errors	not specified	not specified	not specified
	0.25	0.5	1	
	Reading speed errors	reading speed above 16.5 cps	reading speed above 20 cps	reading speed above 25 cps
		0.25	0.5	1

¹ Not specified by Pedersen (2017).

² According to Pedersen (2017), errors in this category are not classified as serious.

5.3.2. Quantitative analysis

We complemented qualitative FAR model analyses with additional quantitative analyses related to the degree of text condensation, which is a key indicator of subtitling expertise (Díaz-Cintas & Remael, 2021). We operationalised condensation with three subtitling quality parameters, which we used as dependent variables: condensation rate, number of subtitles and reading speed. The independent variable was group with three levels: professionals using EdList (PROED), professionals using EZTitles (PROEZT), and trainees (TR). To compare the three groups, a one-way ANOVA was conducted for each dependent variable. We also examined a relation between the participants' age and subtitling quality parameters using Pearson's correlation coefficient. All analyses were performed in SPSS Statistics 27.

We acknowledge that the number of participants in our study compromises the reliability and generalisability of our results. Our analyses should therefore be treated as indications of certain trends and as such require more research.

6. Results: FAR model analyses

Contrary to our hypothesis that professional subtitlers would create subtitles of higher quality, operationalised as higher FAR scores, we found little difference between the professionals and trainees (see Table 2 and 3). In fact, trainees achieved slightly higher scores in two of the three FAR areas and higher total scores. The only area in which professionals achieved higher scores was readability. Table 2 presents the scores for each of the three FAR areas and the total score for each participant in the study.

Table 2: Scores for the FAR areas and the total scores for each participant

Participant	Functional equivalence score	Acceptability score	Readability score	Total score
N01	89.74%	99.00%	96.15%	95.09%
N02	86.36%	97.73%	90.15%	91.41%
N03	90.28%	99.31%	95.83%	95.14%
N04	96.67%	98.89%	92.22%	95.93%
N05	77.14%	97.14%	85.71%	86.67%
P02	83.33%	97.73%	96.21%	92.42%
P04	87.84%	99.00%	91.22%	92.79%
P06	94.32%	97.16%	90.91%	94.13%
P07	87.21%	98.84%	89.53%	91.86%
P08	79.17%	94.17%	91.67%	88.33%
P09	93.42%	99.34%	96.71%	96.49%
P10	82.69%	95.19%	97.12%	91.67%
P11	79.00%	97.00%	99.00%	91.67%
P12	79.69%	99.22%	96.09%	91.67%

Table 3: Means for the functional equivalence score, the acceptability score, the readability score and the total score for the three groups of participants

		N	Mean	SD	95% Confidence Interval			
					Lower Bound	Upper Bound	Min.	Max.
Functional equivalence	PROEZT	6	87.64	5.66	81.70	93.57	79.69	94.32
	PROED	3	80.29	2.08	75.11	85.46	79.00	82.69
	TR	5	88.04	7.14	79.17	96.91	77.14	96.67
Acceptability	PROEZT	6	98.55	0.89	97.61	99.48	97.16	99.34
	PROED	3	95.45	1.43	91.89	99.01	94.17	97.00
	TR	5	98.41	0.93	97.26	99.57	97.14	99.31
Readability	PROEZT	6	93.45	3.23	90.06	96.83	89.53	96.71
	PROED	3	95.93	3.81	86.47	105.39	91.67	99.00
	TR	5	92.01	4.33	86.64	97.39	85.71	96.15
Total score	PROEZT	6	93.23	1.82	91.31	95.14	91.67	96.49
	PROED	3	90.56	1.93	85.77	95.35	88.33	91.67
	TR	5	92.85	3.87	88.04	97.66	86.67	95.93

6.1. Functional equivalence

Overall, out of the three FAR areas, functional equivalence is where both professional and trainee subtitlers achieved the lowest mean score (see Table 3). A one-way ANOVA with functional equivalence score as the dependent variable did not find a significant difference between the groups, ($F(2,1) = 1.980, p = .184, \eta = .265$). Descriptive statistics show that the PROED participants had the lowest overall score, and the highest score was achieved by TR. The highest individual score of 96.67% was achieved by N04, whereas the lowest score of 77.14% was achieved by N05, showing a lack of uniformity in the TR group. These results can be considered rather surprising, as professional subtitlers would be expected to achieve higher scores in all areas because of their expertise, which trainees could still lack or have to a lower degree. The higher score achieved by TR compared to professionals needs to be taken with caution, however, given the small number of participants who took part in this study.

The most frequent type of errors made by the participants were semantic errors. One example of such error was a misunderstanding to whom the name “Mac” refers. Six subtitlers wrongly assumed that it refers to the man in the scene instead of the woman, even though participants were provided with a sheet containing a short description of the characters with their names and pictures at the beginning of the experiment. Interestingly, no trainee subtitler made this mistake.

Another frequent semantic error was omitting, shortening or substituting the name of the Gulf of Mexico, which is pertinent to the plot. Nine participants, professionals and trainees alike, did not retain the full name and replaced it with the words *Meksyk* (“Mexico”), *zatoka* (“the gulf”) or *Atlantyck* (“the Atlantic”), or fully omitted the location of the explosion. In the same subtitle, two participants (one trainee and one professional), decided to translate the phrase “an oil well exploded” as *mamy wyciek* (“we have a spill”) or simply *wyciek* (“a spill”), which conveys an important aspect of the event, but is not the most fortunate choice because an oil spill does not necessarily imply casualties, whereas a more extreme event, such as an explosion, does.

Another problematic element was “I’ll fill you in at the 6:00 rundown”. In this context, “rundown” refers to a meeting of journalists. Some participants incorrectly rendered this utterance as *Napiszę o tym w raporcie* (“I’ll write about it in the report”) or *Wpiszę cię w program o 6.00* (“I’ll put you into the program for 6:00”). Some subtitlers translated it as *Więcej o szóstej* (“More at six”), which is less misleading, but it still does not fully render the meaning of the original.

The next problematic element was the response to “Oh, blow me”, i.e. “I want you to not use that language in front of women and to forever not suggest that image to me”. Although the participants usually succeeded at conveying the meaning of the first utterance, sometimes their translation of the response did not make sense, e.g. *Wal się* (“Go screw yourself”) and *Nie podsuwaj mi takich obrazów* (“Don’t suggest such images to me”) or *Ja pierdolę* (an expletive which takes the first-person singular form) and *Nie sugeruj mi takich rzeczy* (“Don’t suggest such things to me”). Such translations can be considered incorrect because they cause a loss of meaning and they are not a logical reply to the first utterance. Again, a similar number of professionals and trainees made this error.

Regarding measurements, most subtitlers decided to convert feet into kilometres or metres and gallons into litres. In most cases, the conversion was conducted correctly, but two participants – N02 and N05 – used UK gallons instead of US gallons, which affected the conversion. One participant (N02) made a mistake in the conversion of feet, which resulted in the number being visibly too high, i.e., *55 tys. km* instead of *5,5 tys. km*, resulting in an absurdity, as it is impossible to drill this deep into the Earth.

When it comes to the utterance about the name of the oil well, “Deepwater Horizon is aptly named”, the challenge was to translate it in a way which would be understandable to the Polish viewer. The majority of subtitlers used the transfer strategy (Pedersen 2011), that is retained the original term in the target text, but some omitted the entire utterance. Some participants transferred the name while omitting the comment, probably because of temporal and spatial constraints. There were approximately three seconds to display this subtitle, so it would probably be possible to include both the name of the oil well and the comment. A few of the participants also decided to translate the name using dictionary equivalents into *Głęboki Horyzont* or *Horyzont Głębki*. This, although makes the commentary clearer, obscures the original reference, thus causing some loss of meaning.

Overall, participants made fewer stylistic errors than semantic errors: the average number of penalty points for semantic errors was 4.04, whereas that for stylistic errors – 0.57. The most frequent stylistic error was the incorrect choice of terms of address when two employees were addressing their boss: “I need one of your staffers” and “I write your blog”. Some subtitlers decided to use the informal ‘you’ form, whereas it would be more appropriate to use the formal pronoun *pan* because of the difference in age and status and the fact that the characters do not know each other well.

6.2. Acceptability

To determine the effect of expertise on acceptability, we conducted a one-way ANOVA and found a significant effect of group ($F(2,11) = 10.368, p = .003, \eta = .653$) on this FAR parameter. The lowest acceptability score was attained by the PROED group. Post-hoc Tukey test showed that subtitlers using EdList differed significantly from both professionals ($p = .003$) and trainees ($p = .006$) who worked with EZTitles. There was no significant difference between professionals and trainees using EZTitles.

Out of the three FAR areas, the participants made the lowest number of acceptability errors. Although grammar and spelling errors did occur, they were rather rare. It appears that achieving idiomaticity was not a problem for the subtitlers in this experiment, but some participants struggled to render the already-mentioned expression “Oh, blow me”. It seems that the problem stemmed from the desire to translate it in a way which would make sense in the context of the following response. The majority of subtitlers managed to translate this expression in an idiomatic way. Only four participants – two trainees and two professionals – did not render it in a natural way. Some of the translations do not sound idiomatic in Polish, e.g.: *Ssij mi, Pieprz mnie, Pieprzę cię* and *Kierwa*, which is an uncommon euphemism for the swearword.

6.3. Readability

A one-way ANOVA analysis with readability as the dependent variable showed no statistically significant differences between the groups, ($F(2,11) = 1.015, p = .394, \eta = .156$). Contrary to the other two FAR areas, the mean readability score of the professionals was slightly higher than that of the trainees. The highest score in this category was achieved by PROED participants.

All participants made some segmentation errors, although they were usually minor, which means that they were done on the level of line breaks rather than segmentation between subtitles. The most common error made by participants was not placing questions and answers in the same subtitle.

Errors in the punctuation and graphics category were rare and included missing commas and full stops at the end of a sentence. One participant erroneously placed speaker dashes in subtitles with only one speaker in two places. Most errors in this category were classified as minor because they usually did not affect the understanding of the subtitle.

Reading speed errors were made by most of the participants: only participants P08, P11 and P12 did not receive any penalty points for this area. Errors were usually classified as minor, which means that they did not exceed the maximum reading speed by a large margin. The fact that errors of this type were commonly made by the subtitlers in this study probably stems from the fast-paced dialogue in the clip from *The Newsroom*.

6.4. Total scores

Calculating the total FAR score provides a holistic view of the quality of the subtitles created by different participants. It takes into consideration all three areas at once and allows for the comparison between different participants. Total scores can be considered a reflection of the overall subtitling quality achieved by the participants.

All in all, it seems rather surprising that the professional subtitlers did not achieve higher FAR scores than the trainees. Our initial expectation was that expertise would give professionals an advantage, which would translate into higher scores. It is, however, worth noting that the total scores achieved by professionals were more uniform than those achieved by trainees, as shown in Fig. 1. This could suggest that the subtitling quality levels were more consistent among professional subtitlers than among trainee subtitlers.

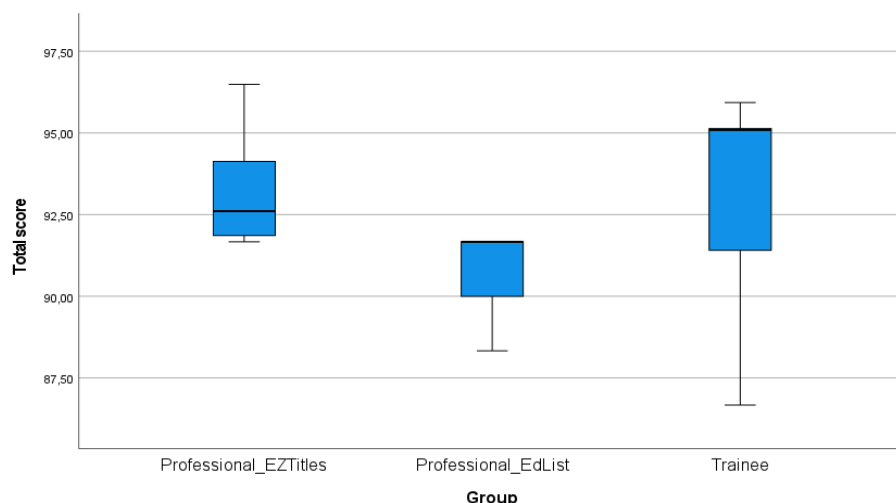


Figure 1: Total score by expertise

7. Results: quantitative analysis

To complement the analysis using the FAR model with quantitative data, we decided to follow it up with two further analyses. First, we compared three key parameters indicative of the condensation in subtitling: condensation rate, number of subtitles and reading speed. Then, we examined the age of the participants, which we thought may constitute a confounding factor with the software used (see section 7.2).

7.1. Condensation

7.1.1. Condensation rate

Condensation rate was calculated as the percentage of words of the target-language subtitles relative to the number of words in the source text (358 words):

$$\text{Condensation rate} = 100\% - \left(\frac{\text{number of words in the target text}}{358} \times 100\% \right)$$

As shown by descriptive statistics in Table 4, condensation rates were highest among professionals working with EdList, who removed more than 55% words compared to the original, and lowest in the case of trainees. Professionals using EZTitles achieved a higher condensation rate than trainees, but lower than professionals using EdList.

Table 4: Subtitling quality parameters per group

	Condensation rate (SD)	Number of subtitles (SD)	Reading speed (SD)
PROEZT	43.97 (2.27)	37.83 (4.95)	15.53 (.71)
PROED	55.65 (1.41)	27.00 (2.64)	13.09 (.07)
TR	40.87 (2.10)	37.60 (4.66)	15.66 (.29)

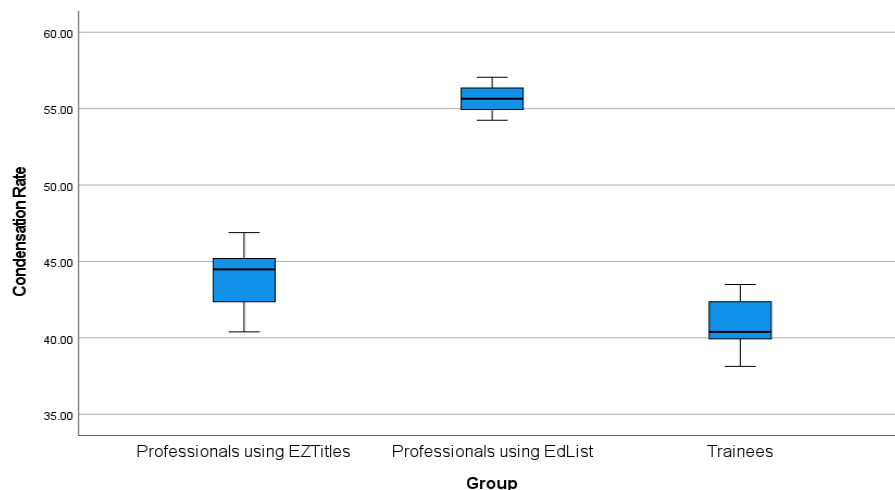


Figure 2: Condensation rate by group

To compare the effect of group (PROED, PROEZT and TR) on the condensation rate, we performed a one-way ANOVA, which showed a statistically significant difference, $F(2, 11) = 49.677, p < .001, \eta = .900$. Post-hoc Tukey's HSD test revealed that the difference was between PROED and PROEZT ($p < .001$), as well as between PROED and TR ($p < .001$). No statistically significant difference was found between professionals using PROEZT and TR ($p = 0.75$).

7.1.2. Number of subtitles

The number of subtitles can be considered a direct result of condensation, since it can be assumed that the fewer subtitles there are, the more the volume of the original dialogue has been reduced. This measure is correlated with the condensation rate measured in the number of words discussed above.

Out of the three groups, PROED created the lowest number of subtitles, whereas PROEZT and TR achieved very similar results. A one-way ANOVA showed a significant difference between at least two groups, $F(2, 11) = 6.663, p = .013, \eta = .548$. Tukey's test indicated that there was a significant difference between PROED and PROEZT ($p = 0.15$) and between PROED and TR ($p = .021$). There was no significant difference between PROEZT and TR.

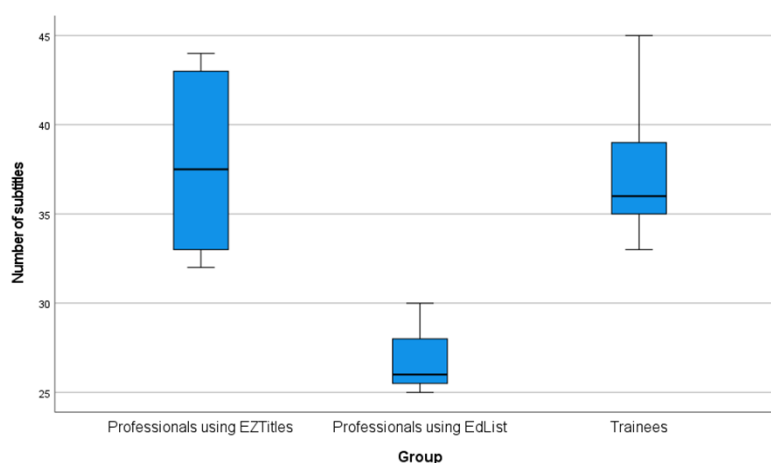


Figure 3: Number of subtitles by group

7.1.3. Reading speed

The mean reading speed was calculated using the third approach proposed by Fresno and Sepielak (2020), which seems to be the most successful at reflecting that each subtitle has its own speed, rather than simply treating the subtitle set as a uniform whole. The formula used for calculating mean reading speed consists in adding the reading speeds of each subtitle and dividing the sum by the number of subtitles:

$$\text{Mean reading speed} = \frac{RS1 + RS2 + + RSn}{n}$$

PROED had the lowest mean reading speed out of the three groups (see Figure 4). The other two groups had very similar speeds, although PROEZT had slightly lower speeds than TR. To compare the effect of group on reading speed, a one-way ANOVA was performed. It revealed that there was a statistically significant difference between at least two groups, $F(2, 11) = 28.214, p < .001, \eta = .674$. Tukey's test showed that, again, there was a difference between PROED and PROEZT ($p < .001$) and between PROED and TR ($p < .001$). There was no significant difference between PROEZT and TR.

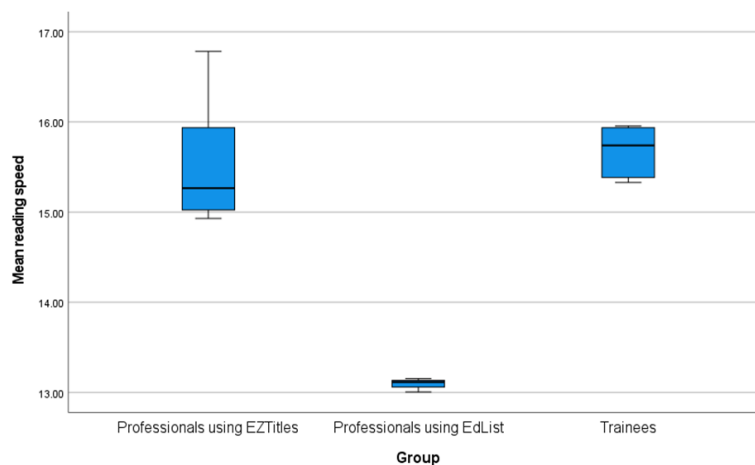


Figure 4: Reading speed by group

Overall, PROEZT and TR created a similar number of subtitles with similar condensation rates and reading speeds, whereas PROED created significantly fewer subtitles which were more condensed and had lower reading speeds. Apart from expertise, this result may also be potentially attributed to differences in subtitling tools: at the time when the study was conducted, only EZTitles allowed to set the reading speed in cps, whereas EdList had a built-in error indicator which could not be customised.

As Orrego-Carmona et al. (2018) point out, this similarity between PROEZT and TR might also be due to the fact that participants in these groups were closer in age than those who used EdList. PROED were older participants and therefore they were probably used to working with a lower reading speed and a lower maximum number of characters per line for a significant portion of their careers (Orrego-Carmona et al., 2018, cf. Ivarsson & Carroll, 1998). By contrast, most of the participants in both groups using EZTitles were younger, which means that they have probably become accustomed to higher reading speeds and maximum numbers of characters, which are becoming more and more widespread (cf. Netflix's 42 characters per line and 17 cps).

7.2. Participants' age

It can be assumed that a subtitler's age can be linked to their experience, since professionals who have been in the job for longer are more likely to have more

experience of deliberate practice (Tiselius and Hild 2017, p. 434) than younger ones. To determine whether there was a relationship between the participants' age and the possible indications of experience – i.e. condensation rate, number of subtitles, and reading speed – the Pearson correlation coefficient was calculated.

Table 5: Correlations between participants' age, condensation rate, reading speed, and number of subtitles

		Age	Condensation Rate	Mean reading speed	No. of subtitles
Age	Pearson Correlation	--			
	N	14			
Condensation Rate	Pearson Correlation	.784**	--		
	Sig. (2-tailed)	.001			
	N	14	14		
Mean reading speed	Pearson Correlation	-.749**	-.957**	--	
	Sig. (2-tailed)	.002	<.001		
	N	14	14	14	
Number of subtitles	Pearson Correlation	-.552*	-.779**	.737**	--
	Sig. (2-tailed)	.041	.001	.003	
	N	14	14	14	14

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

We found a positive correlation between the participants' age and the condensation rate, $r(12) = .784, p < .001$. There was also a negative correlation between participants' age and reading speed, $r(12) = -.749, p = .002$, and between participants' age and number of subtitles, $r(12) = -.552, p = .041$. This shows that the older the participants were, the higher the condensation rate they achieved, the lower the number of subtitles they created and the lower the reading speed of these subtitles was. This may be because professionals using EdList may have different subtitling habits given that they started their careers when different standards were in place.

8. Discussion and conclusions

The main aim of this study was to explore the relationship between the quality of interlingual subtitles and subtitling expertise. With this aim in mind, we

compared the quality of subtitles created by trainees and professionals. We conducted two sets of analyses: a qualitative analysis using Pedersen's (2017) FAR model and a quantitative analysis of subtitling quality parameters related to text condensation. We expected that professionals would achieve higher subtitling quality scores in the three FAR model areas and that they would have higher condensation rates compared to trainees. Our professionals were recruited on the understanding that audiovisual translation tasks, including interlingual subtitling, were their main source of income. Our trainees, on the other hand, were all MA translation students who had completed an optional 30-hour subtitling course as part of their studies.

Contrary to our hypotheses, professional subtitlers did not achieve higher scores than trainees in the FAR model analyses. What is more, the subtitles created by the most experienced professionals were not necessarily error-free: professional subtitlers and trainees alike made different types of errors, including the most serious ones such as semantic or idiomaticity errors. In other words, contrary to our predictions, we did not find a relationship between expertise and subtitling quality as measured by the FAR model.

One of the questions we asked at the beginning of this paper relates to what distinguishes a subtitling professional from a novice. The results of our study indicate that a key difference may lie in condensation skills. In our study, the subtitlers with the longest professional experience created subtitles that were the most condensed, as indicated by the highest condensation rate (over 50% relative to the original), the smallest number of subtitles and the lowest reading speeds. These three subtitling quality parameters go hand in hand with one another, and becoming an expert in subtitling requires mastering them all. These results support the understanding of expertise as the supreme expression of professional competence, as postulated by Tiselius and Hild (2017, p. 425). Condensation skills are a top priority in subtitling, as reiterated by numerous subtitling tutors, style guides, textbooks as well as professional subtitlers themselves (Belczyk, 2007; Díaz-Cintas & Remael, 2021; Pedersen, 2011; Szarkowska, 2016).

Another interpretation of these results is related to the fact that subtitling conventions are changing: while low reading speeds (of around 10-12 cps) and high condensation rates were common two or three decades ago (Ivarsson & Carroll 1998), the rise in reading speeds has resulted in loosening the requirements related to condensation in accordance with the rule that the higher the reading speeds allowed, the lesser degree of text condensation is required. This shift in reading speeds triggers changes in professional practices. Newcomers to the field are not expected to condense text to the extent that previous trainees did. Nevertheless, those accustomed to traditional subtitling methods persist in their practices, highlighting a sense of adherence to established norms. This suggests a resistance to change and underscores the significant role of familiar conventions in shaping one's approach to subtitling.

Our study has also contributed to evaluating the usefulness of the FAR model in subtitling quality assessment. The largest drawback of the FAR model includes the subjectivity of penalty points and their severity – something which Pedersen (2017) admits. However, removing subjectivity from assessment would prove challenging, if not impossible. Some errors are difficult to classify into a category, for example not spelling out numbers from zero to ten could potentially be classified as a stylistic or a spelling error. Furthermore, Pedersen's model does not take into consideration timing issues such as shot changes or chaining, but these error categories could probably be added by the assessor if necessary. On the positive side, the FAR model allows for a comprehensive assessment between subtitlers, whose scores can be directly

compared. The model – though subjective and time-consuming to implement – can also be considered easy to apply in practice and to customise.

Finally, we acknowledge some limitations of this study, including the low number of participants, the fact that they came from only one country and had only one language combination (English-Polish) as well as the fact that subtitlers used two different types of software and the lab setting in which the subtitling task was conducted. The low number of participants may have affected the results, especially when professionals using EdList were considered. Therefore, the results presented will need to be confirmed through more studies with more participants that provide additional evidence.

References

- ARTE. (2019). *Complete technical guidelines ARTE G.E.I.E.* <https://www.arte.tv/sites/corporate/files/complete-technical-guidelines-arte-geie-v1-07-2.pdf>
- Baker, J., Hodges, N., & Wilson, M. (2018). Collecting and assessing practice activity data: Concurrent, retrospective, and longitudinal approaches. In K. Ericsson, R. Hoffman, A. Kozbelt, & A. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 257-270). Cambridge University Press. <https://doi.org/10.1017/9781316480748.015>
- Belczyk, A. (2007). *Tłumaczenie filmów* [Translating films]. Wydawnictwo Dla Szkoły.
- Beuchert, K. (2017). *The web of subtitling: A subtitling process model based on a mixed methods study of the Danish subtitling industry and the subtitling processes of five Danish subtitlers* [Doctoral dissertation, Aarhus University], https://pure.au.dk/portal/files/119118393/PhD_KathrineBeuchert.pdf.
- Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment. Machine translation: Technologies and applications* (pp. 9-38). Springer. https://doi.org/10.1007/978-3-319-91241-7_2
- Channel 4. (n.d.). *Channel 4 subtitling guidelines.* https://www.channel4.com/media/documents/corporate/foi-docs/SG_FLP.pdf
- Chmiel, A. (2021). Effects of simultaneous interpreting experience and training on anticipation, as measured by word-translation latencies. *Interpreting*, 23(1), 18–44. <https://doi.org/10.1075/intp.00048.chm>
- Christoffels, Ingrid K., Annette M. B. de Groot, and Judith F. Kroll. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, 54(3), 324-345. <https://doi.org/10.1016/j.jml.2005.12.004>.
- Díaz-Cintas, J., & Remael, A. (2014). *Audiovisual translation: Subtitling.* Routledge. <https://doi.org/10.4324/9781315759678>
- EMT. (2017). European Master's in Translation. Competence Framework 2017. https://commission.europa.eu/system/files/2018-02/emt_competence_fw_2017_en_web.pdf
- ESIST. (2021). *AVT Guidelines and Policies.* <https://www.esist.org/resources/avt-guidelines-and-policies/>
- Ericsson, K.A., Charness, N., Feltovich, P.J. & Hoffman, R.R. (2006). *The Cambridge handbook of expertise and expert performance.* Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796>
- Ericsson, K. (2018). Superior working memory in experts. In K. Ericsson, R. Hoffman, A. Kozbelt, & A. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (Cambridge Handbooks in Psychology, pp. 696-713). Cambridge University Press. <https://doi.org/10.1017/9781316480748.036>
- Fresno, N., & Sepielak, K. (2020). Subtitling speed in Media Accessibility research: Some methodological considerations. *Perspectives*, 30(3), 415-431. <https://doi.org/10.1080/0907676X.2020.1761841>
- Göpferich, S. (2009). Towards a model of translation competence and its acquisition: The longitudinal study 'TransComp'. In S. Göpferich, A. L. Jakobsen, & I. Mees

- (Eds.), *Behind the mind: Methods, models and results in translation process research*. Samfundslitteratur.
- Göpferich, S. (2013). Translation competence: Explaining development and stagnation from a dynamic systems perspective. *Target. International Journal of Translation Studies*, 25(1), 61-76. <https://doi.org/10.1075/target.25.1.06goe>
- Hurtado Albir, A. (2017). Translation and translation competence. In A. Hurtado Albir (Ed.), *Researching translation competence by PACTE Group* (pp. 3-34). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.127>
- Ivarsson, J., & Carroll, M. (1998). *Code of Good Subtitling Practice*. <https://www.esist.org/wp-content/uploads/2016/06/Code-of-Good-Subtitling-Practice.PDF.pdf>
- Künzli, A. (2021). From inconspicuousness to flow—The CIA model of subtitle quality. *Perspectives*, 29(3), 326–338. <https://doi.org/10.1080/0907676X.2020.1733628>
- Namkung, V. (2021). Translators, experts weigh in on ‘Squid Game’ subtitle debate. *NBC News*. <https://www.nbcnews.com/news/asian-america/translators-experts-weigh-squid-game-subtitle-debate-rcna2568>
- Netflix. (n.d.-a). *Backlot automated QC error messages – Netflix | Partner Help Center*. <https://partnerhelp.netflixstudios.com/hc/en-us/sections/203547178-Backlot-Automated-QC-Error-Messages>
- Netflix. (n.d.-b). *Introduction to Netflix Quality Control (QC)*. Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/115000353211-Introduction-to-Netflix-Quality-Control-QC>
- Netflix. (n.d.-c). *Timed text—manual QC error messages – Netflix | Partner Help Center*. <https://partnerhelp.netflixstudios.com/hc/en-us/sections/360000285407-Timed-Text-Manual-QC-Error-Messages>
- Netflix. (2021a). *English timed text style guide*. Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/217350977-English-Timed-Text-Style-Guide>
- Netflix. (2021b). *Polish timed text style guide*. Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/216787928-Polish-Timed-Text-Style-Guide>
- Netflix. (2021c). *Timed text style guide: General requirements*. Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215758617-Timed-Text-Style-Guide-General-Requirements>
- Netflix. (2021d). *Timed text style guide: Subtitle timing guidelines*. Netflix | Partner Help Center. <https://partnerhelp.netflixstudios.com/hc/en-us/articles/360051554394-Timed-Text-Style-Guide-Subtitle-Timing-Guidelines>
- Orrego-Carmona, D., Dutka, L., & Szarkowska, A. (2018). Using translation process research to explore the creation of subtitles: An eye-tracking study comparing professional and trainee subtitlers. *JoSTrans: The Journal of Specialised Translation*, 30, 150-180.
- Pedersen, J. (2011). *Subtitling norms for television: An exploration focussing on extralinguistic cultural references*. John Benjamins.
- Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *Journal of Specialised Translation*, 28, 210-229.
- Pedersen, J. (2018). From old tricks to Netflix: How local are interlingual subtitling norms for streamed television? *Journal of Audiovisual Translation*, 1(1), 81-100. <https://doi.org/10.47476/jat.v1i1.46>
- Robert, I., & Remael, A. (2016). Quality Control in the Subtitling Industry: An Exploratory Survey Study. *Meta*, 61(3), 578–605. <https://doi.org/10.7202/1039220ar>
- Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy rate in live subtitling: The NER model. In R. B. Piñero & J. Díaz-Cintas (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Palgrave Macmillan UK. https://doi.org/10.1057/9781137552891_3
- Romero-Fresco, P., & Pöschhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR model. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 149–167. <https://doi.org/10.52034/lanstts.v16i0.438>
- Szarkowska, A. (2016). *Report on the results of an online survey on subtitle presentation times and line breaks in interlingual subtitling. Part 1: Subtitlers*.

- University College London. http://avt.ils.uw.edu.pl/files/2016/10/SURE_Report_Survey1.pdf
- Szarkowska, A., Díaz-Cintas, J., & Gerber-Morón, O. (2021). Quality is in the eye of the stakeholders: What do professional subtitlers and viewers think about subtitling? *Universal Access in the Information Society*, 20(4), 661–675. <https://doi.org/10.1007/s10209-020-00739-2>
- Whyatt, B. (2018). Testing indicators of translation expertise in an intralingual task. *Hermes (Denmark)*, 57, 63-78. <https://doi.org/10.7146/hjlcb.v0i57.106194>
- Tardel, A. (2020). Effort in semi-automatized subtitling processes: Speech recognition and experience during transcription. *Journal of Audiovisual Translation*, 3(2), 79–102. <https://doi.org/10.47476/jat.v3i2.2020.131>
- TED. (2017). *English Style Guide*. http://translations.ted.org/wiki/English_Style_Guide
- Tiselius, E., & Hild, A. (2017). Expertise and competence in translation and interpreting. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 423-444). Wiley. <http://dx.doi.org/10.1002/9781119241485.ch23>