QUALITY OF POST-EDITED INTERLINGUAL SUBTITLING: FAR MODEL, TRANSLATOR'S ASSESSMENT AND AUDIENCE RECEPTION

Arlene Koglin[1, 2]
[1]Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil
[2]Universidade Federal de de Santa Catarina, Florianópolis, Santa Catarina, Brasil

João Gabriel Pereira da Silveira[1]
[1]Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil

Morgana Aparecida de Matos[2, 3]
[2]Universidade Federal de de Santa Catarina, Florianópolis, Santa Catarina, Brasil
[3]Universidade de Vigo, Vigo, Galicia, Espanha

Vitória Tassara Costa Silva[2]
[2]Universidade Federal de de Santa Catarina, Florianópolis, Santa Catarina, Brasil

Willian Henrique Cândido Moura[2]
[2]Universidade Federal de de Santa Catarina, Florianópolis, Santa Catarina, Brasil

Abstract: This paper analyzes the quality of machine-translated interlingual subtitles, which were post-edited in the language pair EN/PT-BR. Our analyses applied the FAR model, a Translation Quality Assessment Model, to the PT-BR subtitles of *The Red Sea Diving Resort* movie trailer, correlating it to empirical data collected with translators (quality assessment) and audience (reception). Reception data was collected with undergraduate students, which were divided into two groups: the control group that watched the subtitled trailer available on Netflix; and the experimental group that watched the trailer with post-edited subtitles. Quality data was collected with translators and they watched the trailer with post-edited subtitles. We used a 5-point Likert-

type scale, a questionnaire and a guided think-aloud protocol to collect our data. Data collected with both translators and students were correlated with the FAR model error scores (Functional Equivalence, Acceptability and Readability). Our results indicate that the post-edited subtitles had good quality in terms of meaning and target language norms; however, the technical parameters had lower quality, which affected the trailer appreciation as reported by most of the audience. Due to the small sample size, further empirical studies are required to obtain solid standards for assessing the quality of post-edited subtitles.

Keywords: Subtitling Quality Assessment; Reception; Machine Translation; Post-Editing; FAR Model

QUALIDADE DE LEGENDAS INTERLINGUAIS PÓS-EDITADAS: FAR MODEL, AVALIAÇÃO DE TRADUTORES E RECEPÇÃO DA AUDIÊNCIA

Resumo: Este artigo analisa a qualidade de legendas interlinguais traduzidas automaticamente e pós-editadas no par linguístico EN/PT-BR. A análise aplicou o FAR model, um Modelo de Avaliação da Qualidade de Tradução, às legendas em PT-BR do trailer do filme *The Red Sea Diving Resort,* no Brasil: *Missão no Mar Vermelho*, correlacionando-o com dados empíricos coletados com tradutores (avaliação da qualidade) e com a audiência (recepção). Os dados de recepção foram coletados com alunos da graduação, que foram divididos em dois grupos: o grupo controle, que assistiu ao trailer com as legendas disponibilizadas pela Netflix; e o grupo experimental, que assistiu ao trailer com as legendas pós-editadas. Os dados de avaliação da qualidade foram coletados com tradutores que assistiram ao trailer com as legendas pós-editadas. Os instrumentos de coleta foram uma escala Likert de 5 pontos, um questionário e protocolos verbais guiados. Os dados coletados com os participantes foram correlacionados com as categorizações de erros do FAR model (Equivalência Funcional, Aceitabilidade e Leiturabilidade). Os resultados indicaram que as legendas pós-editadas possuem qualidade boa em termos de significado e normas da língua-alvo; entretanto, os parâmetros técnicos tiveram qualidade inferior, o que afetou a apreciação do trailer conforme relatado pela maioria da audiência. Devido à amostra reduzida, são necessários estudos empíricos adicionais para a obtenção de padrões mais sólidos de avaliação da qualidade de legendas pós-editadas.

Palavras-chave: Avaliação da Qualidade de Legendagem; Recepção; Tradução Automática; Pós-Edição; FAR Model

1. Introduction

Quality assessment of interlingual subtitling is a promising field of empirical studies with many aspects that still lack investigation. One of them is related to the quality of machine-translated subtitles. As a modality of Audiovisual Translation (AVT), subtitling involves the translation of original dialogues and other verbal information into a written text on the target language, which appears on the screen (Díaz Cintas, 2012, p. 274). Subtitling has consolidated itself as one of the most popular and in-demand practices of AVT and has been following the technological advances of the last decades. Consequently, the translation and subtitle generation process has been optimized due to the number of translation software options and subtitle generation tools.

In turn, machine translation (MT) happens to optimize the subtitling process as it consists of transferring to the computer the task of translating texts from one language to another. Therefore, the translation improvement process is not only in terms of time and effort, but mainly in terms of maintaining a high level of terminological consistency (Athanasiadi, 2017, p. 31). In the audiovisual industry, subtitling is considered one of the most expensive tasks because, like other types of AVT, it needs to be performed by a specialist due to the linguistic and technical specificities that it demands. In this context, machine translation and subtitling can form an intersection since the MT can help increase productivity and quality, where human translation appears in the post-editing process, focusing on quality and specific linguistic approaches to develop the subtitle production.

The GETRADTEC Group from the Federal University of Pernambuco, Brazil, has been developing a project within this scope through an empirical-experimental approach. GETRADTEC project aims to investigate the reception and the probable effects on quality, technical parameters, and linguistic aspects of machine-translated interlingual subtitles. Therefore, this paper presents the results of two pilot studies conducted

by the GETRADTEC Group, whose data are preliminary and relevant for further studies in the project.

More specifically, this article aims to analyze the quality of machine-translated interlingual subtitles post-edited by humans, by applying the FAR model to the subtitles of a movie trailer and correlating the FAR model results with the translator's quality assessment and audience reception of the same movie trailer.

The FAR model, developed by Pedersen (2017), has been chosen as the fundamental axis for analyzing the quality of subtitles. Based on error analysis, it consists of a generalized model applied to subtitling, focusing on evaluating the final product and encompasses Functional Equivalence, Acceptability, and Readability of subtitles (Pedersen, 2017, p. 218-224).

The article is divided into five major sections. Firstly, we present in section 2 the theoretical framework, which encompasses a discussion about quality assessment in Translation Studies and in subtitling. Next, we present some quality assessment models with a focus on the FAR model. Then, section 3 explains the methodological aspects regarding data collection and data analysis. In section 4, data is analyzed and discussed. Finally, section 5 brings our final remarks about the findings as well as the limitations of the research.

2. Theoretical Framework

This section presents the theoretical framework of this paper. We will briefly discuss the main concepts of quality assessment in Translation Studies, more specifically quality assessment in subtitling. Next, some quality assessment models will be discussed with focus on the FAR model.

2.1 Quality Assessment in Translation Studies

Assessing the quality of a product is a complex phenomenon and the concept of quality itself can carry several meanings depending

on its approach (Pedersen, 2017), which causes hesitation in scholars regarding the definition of the concept in various areas of knowledge. In Translation Studies, the concept of quality is challenging because it involves the subjectivity of individuals and value judgments motivated by individual reasons. Although it is a difficult task to consider such types of judgments in order to achieve scientific objectivity, this research area should not be seen as a worthless one (House, 2001, p. 255).

In the translation industry, translation quality assessment (TQA) is closely linked to translation management and translation process, and there is a considerable amount of research that provides an applied perspective of TQA in this context (cf. Doherty, 2017; Doherty *et al.*, 2013; Gaspari, Almaghout & Doherty, 2014; Pym, 2020). The standards established by ISO EN 17100:2015, issued by the International Organization for Standardization, and EN 15038:2006, from the European Committee for Standardization, for instance, specify the competences and qualities of professionals in charge of the translation process, such as translators, reviewers, terminologists and managers, who are closely linked to quality assurance with a focus on suppliers and their respective customers (Szarkowska, Díaz Cintas & Gerber-Morón, 2020, p. 2). In addition to that, deductive assessment models based on counting errors and applying penalties based on the severity of errors are also used, especially in the localization industry and IT (Doherty, 2017; O'Brien, 2012).

In the academic field, TQA tends to be viewed from a communicative perspective and the focus falls on equivalence issues, with quality generally being assessed through translation models categorized into three main prisms: response-based approaches, text-based approaches, and functional-pragmatic approaches (Szarkowska, Díaz Cintas & Gerber-Morón, 2020, p. 3).

Response-based approaches have the equivalence between the translation and the original text as the main focus of investigation from the point of view of their respective target audiences in order to assess whether the response provided by the consumer audience of the translation is equivalent to the responses provided by the

consumer audience of the original work (House, 2005). This type of approach was theorized under the dynamic equivalence, a concept proposed by Nida (1964), who postulated the notions of "informativeness" and "intelligibility" as the main criteria to evaluate the quality of a translation, and from the perspective of Gutt's (2014) relevance-theoretic model.

Text-based approaches, which are largely rooted in Linguistics, emphasize the comparison of the source text (ST) with the target text (TT) in order to identify the main strategies used by the translator in terms of syntactic, stylistic and semantic changes (Szarkowska; Díaz Cintas & Gerber-Morón, 2020, p. 3) and can also be discussed from different theoretical perspectives. From the perspective of Comparative Literature, for instance, the quality of a translation is assessed according to the form and function of the translation in the cultural and literary system of the TT (cf. Toury, 1995), while from the point of view of Functionalist Theory, the focus is on the Skopos, i.e., on the purpose of the translation (cf. Reiss & Vermeer, 1984).

Lastly, we can mention the functional-pragmatic approach, which seeks to evaluate quality based on the pragmatic perspectives of language use. Supported by Halliday's systemic-functional theory, House (2001, 2005) has developed a translation evaluation model based on the analysis of ST and TT segments, in order to make comparisons and evaluations according to the relative correspondence between them, establishing as a basic requirement for equivalence the presence of a function of the TT that is equivalent to the ST. According to Abdelaal (2019), the author's model was first proposed in 1981, was reviewed in 1997 and more recently in 2015, and its most recent version is applicable to subtitling assessment, which will be treated with more detail in the following subsection (Abdelaal, 2019, p. 7).

2.2 Quality Assessment in Subtitling

Audiovisual Translation (AVT) has established itself as a relevant area for Translation Studies and a considerable amount

of research has been conducted in various institutions across the world, especially regarding subtitling. According to Gottlieb (2005), subtitling is a translation modality that involves the overlap of a written text on the screen synchronized with the verbal text of the audiovisual product. In this modality, "the speech act is always in focus; intentions and effects are more important than isolated lexical elements" (Gottlieb, 2005, p. 247), and there is also a series of technical parameters (space, number of lines, characters per line, characters per second) that need to be respected by the translator so that the subtitles convey the ST's message consistently.

The current Brazilian audiovisual context can be considered highly heterogeneous, since consumers from diverse profiles consume different kinds of audiovisual productions, national and foreign, through different platforms and settings. In terms of the consumption of audiovisual products, Brazil has a tradition of being a country that avidly consumes foreign audiovisual material (Alfaro de Carvalho, 2012, p. 468), translated into their respective modalities – dubbing, subtitling, voice-over, Closed Captions, etc. – according to the specificities of the materials and the setting – Cinema, open TV, cable TV, streaming platforms, etc.

Considering the scope of this paper, that is, interlingual subtitled productions, historically we can observe that the cinema and cable TV played an important role on introducing subtitled material into Brazilian Portuese. The majority of this material is originated from the United States (Alfaro de Carvalho, 2012, p. 468) and has English as the source language. Currently, the Brazilian scenario presents a new variable: the movement of streaming platforms[1],

---

[1] Content distribution process, via the Internet, in which the user begins viewing files without having to download them, allowing quicker viewing with the content displayed sequentially, as it arrives at the user's computer. The user will be viewing the contents of the files at the rate they arrive, requiring only a small initial waiting time for the synchronization process and the creation of a temporary memory (buffer) used to store a few seconds of content, to absorb changes in the reception rate and/or temporary connection breaks (Adão, 2006, p. 21, translated by Campos & Azevedo, 2020, p. 225-226).

such as YouTube, Netflix, Amazon Prime Video, Hulu, Disney+, among others, that modernized the access to audiovisual materials in the country, presenting different options of productions (dubbed, subtitled, audio described), originated from many countries and reaching various layers of the Brazilian society.

Given its historical nature concerning its high circulation in Brazil, we can state that subtitling was introduced quite recently to the mass Brazilian public, especially when compared to other translation modalities, such as literary translation. Taking that into consideration, some questions arise concerning the issue of what would be an ideal subtitle for this heterogeneous audience and how could the quality of the subtitles be assessed.

Furthermore, Brazil is a continental country, which only makes it more challenging to obtain a standardized rule to what is considered a good subtitle. In addition, the profile of the consumers and the services used by them to watch the productions vary drastically. For example, in some "online video hosting sites like YouTube and Vimeo […] we can now find a new generation of users who exhibit different viewing behavior" (Rabêlo, Garcia-Murillo & Couto, 2017, p. 483). Regarding streaming platforms, it is possible to observe a great deal of Brazilian consumers of subtitled products obtaining them on streaming platforms, such as Netflix[2], which in one way or another, ends up setting a standard for the quality of subtitles in the country.

This is to say that many variables need to be considered when discussing the quality assessment of interlingual subtitles in Brazil, not to mention the importance of conducting more empirical research about the reception of subtitled productions in the country. Not only the adequacy to the established technical parameters (line length, characters per second, etc.) are to be considered when assessing subtitling quality, but also some other factors, such as: the audience profile (country, social aspects, age), the genre of the

---

[2] In December 2017, Netflix had six million subscribers from Brazil (Dias & Navarro, 2018, p. 19).

audiovisual production (comedy, art films, documentary), the type of subtitles (professional, fansub), etc. These factors will influence on how parameters are created to evaluate the subtitles within their context of production/reception.

With the growth of subtitling as a scientific area, the field has faced several challenges regarding not only the application of technical parameters that have different perspectives in different places of audiovisual consumption, but also regarding the methods that can be used as parameters for assessing the translation quality of interlingual subtitling. Due to the fact that TQA in subtitling started to gain a more relevant status in research institutions, more recent evaluation models such as the NER model (Romero-Fresco & Martínez Pérez, 2015) and the FAR model (cf. Pedersen, 2017) have been developed in an attempt to fill in this gap.

The purpose of these models is to create defined standards for assessing the quality and/or reception of the subtitles based on different types of subtitles (interlingual, intralingual, etc). Considering that "reception studies focusing on interlingual subtitling are a relatively recent phenomenon" (Nikolić, 2018, p. 182), different research methodologies and theoretical approaches, such as the ones focusing on the final product are also worthwhile conducting – combining quantitative and qualitative data, for instance. Among the evaluation models that focus on the product, we have selected the FAR model to assess the quality of interlingual subtitles.

2.3 The FAR Model

Traditionally the word error rate (WER) method has been applied (Romero-Fresco & Martínez Pérez, 2015) to assess the subtitles' quality of audiovisual productions. It consists of dividing the number of errors – from a set of categories – by the total amount of words in the subtitle. Romero-Fresco & Martínez Pérez (2015) affirm that this method was mainly used in the evaluation of live subtitles and it sometimes missed important features of other types of subtitles, which led other scholars to adapt the method into models that

considered other specificities of subtitling, such as the CRIM model, the NERD model, the NER model and the FAR model itself.

The FAR model, adapted from the NER model[3] (Romero-Fresco & Martínez Pérez, 2015) and developed by Jan Pedersen (Stockholm University), is a generalized model designed to evaluate the quality of interlingual subtitles. It can be applied to entire movies, TV programs or just excerpts (Pedersen, 2017, p. 211) and the focus of the evaluation is the final product (the subtitles). The author defines it as a "tripartite" model: the first part evaluates the functional equivalence of the subtitles, the second part evaluates the acceptability (grammaticality, idiomaticity issues, etc.), and the third part seeks to evaluate the readability of the subtitles, which refers to the reading speed, the color of the subtitles, the use of italics and other general technical aspects (Pedersen, 2017, p. 218-224).

The FAR model is viewer-based and based on the premise that a relationship is established between subtitlers and viewers, metaphorically named by the author as a "contract of illusion" (Pedersen, 2007, p. 46-47 *apud* Pedersen 2017, p. 215). This contract is firmed "when viewers pretend that subtitles are the real dialogue, which in fact they are not" (Pedersen, 2017, p. 215) and, in return, the subtitlers "help viewers suspend their disbelief by making their subtitles as unobtrusive as possible" (Pedersen, 2017, p. 215).

Furthermore, Pedersen's model is based on error-analysis and for each identified error a penalty point is assigned, of which the score varies according to the severity of the error, which is subcategorized into minor, standard and serious errors. Minor errors are those that may go unnoticed and only break the illusion contract if viewers are very attentive whereas standard errors are those that tend to break the contract of illusion and ruin the subtitle for most viewers. Serious errors, on the other hand, are those

---

[3] The NER model (acronym for Number of words in the text, Edition errors and Recognition errors) was designed by Romero-Fresco & Martínez Pérez (2015) to evaluate the accuracy of live subtitles. Some of the FAR model›s errors category are derived from it.

errors that not only break the illusion contract, but may also affect the subtitle in which the error is contained as well as the subsequent subtitles, even forcing the viewer to take time and resume reading the subtitles (Pedersen, 2017, p. 217).

Regarding the three main categories that the model brings forward (also the acronym for which FAR stands for), we will now explain each one individually. **Functional Equivalence** is related to the subtitle conveying the message that is meant on the spoken utterance. The concept of equivalence is understood in the model as a pragmatic one, which highlights the importance of combining in the subtitle "both what is said and what is meant" (Pedersen, 2017, p. 218). Equivalence errors are then categorized as semantic and stylistic errors.

Semantic equivalence errors can be categorized as minor (error score 0.5) and include mainly lexical errors, including terminology errors that do not affect the plot. Standard errors (error score: 1.0) stand for a "subtitle that contains errors, but still has bearing on the actual meaning and does not seriously hamper the viewers' progress beyond that single subtitle. Standard semantic errors would also be cases where utterances that are important to the plot are left unsubtitled" (Pedersen, 2017, p. 219). Serious errors (score: 2.0) are the ones that jeopardize the understanding of the subtitle itself and therefore affect the comprehension of the plot and break the contract of illusion.

The second category, **Acceptability**, concerns whether the subtitle sounds foreign or unnatural to the viewer. There are three types of errors in this category: i) grammar errors, ii) spelling errors, and iii) errors of idiomaticity (Pedersen, 2017, p. 220). Grammar errors are bound to the grammar of the subtitle's language. The serious ones cause the subtitles to be difficult to read and/or understand, the minors ones are related to very specific grammatical issues, such as the misuse of whom/who in English. The standard errors are located between these two categories. Minor spelling errors are, for example, missing a letter or other errors that do not jeopardize the overall understanding. Standard spelling errors change the

meaning of the word on the subtitle and the serious ones generate an impossibility to read the word. Idiomaticity errors affect the naturality of the subtitles, that is, they sound unnatural and cause a feeling of strangeness to the native viewers of the subtitles, most of the times being caused by source text interference – "and sometimes this […] interference can become so serious that it becomes an equivalence issue" (Pedersen, 2017, p. 221).

The third category, **Readability**, brings forward some technical issues that may disrupt the comfort of the viewers. It is divided into i) errors of segmentation and spotting, ii) punctuation and iii) reading speed and line length (Pedersen, 2017, p. 222). The first category (i) is set for errors of spotting, that are related to bad synchronization with the speech or the image (delayed or forward), and for segmentation errors, that are related with the break on the semantic or syntactic structure of the subtitle (between one another – more serious, or within the same subtitle – less serious). The second category (ii) concerns the misuse of some features such as italics, dashes, and other types of punctuation and graphics. The severity of errors in this category depends on the guidelines formerly used on the production of the subtitles. The components of the last category (iii), the reading speed and line length, may also depend on the guidelines used to produce the subtitles and the tradition of the country that produced them. The author suggests that in case these guidelines are not accessible, subtitles with a higher reading speed than 15 characters per second (cps) should be penalized (Pedersen, 2017, p. 223-224).

In sum, those are the categories that the FAR model takes into account when analyzing the quality of interlingual subtitles. Some limitations of the model are that it is based on an error-analysis – not leaving space for scoring the good subtitles –, and the subjectivity on judging idiomaticity and equivalence errors (Pedersen, 2017, p. 224). Despite that, this model can be applied to a wide variety of data and subtitled productions, not to mention the advantages of the score penalty in the three categories, which could be helpful to subtitler's training and to giving feedback to translators.

We will now explain how our empirical data was collected and analyzed in this study.

3. Methodology

The use of machine translation to translate subtitles has been studied by researchers in the fields of audiovisual translation and technology in order to assess whether MT can be helpful or not for subtitlers. Developed by the GETRADTEC Group, this piece of research is qualitative and quantitative with a focus on translation as a product. It aims at assessing the quality of subtitles that were machine translated and post-edited. To do so, the FAR model has been applied and its results have been correlated with data collected in two pilot experiments with translators and undergraduate students, as explained in the following subsections.

3.1 Data Collection

3.1.1 Machine Translation and Post-Editing of Subtitles

Initially, all dialogues of the selected trailer were transcribed. Then they were machine translated and post-edited. The software Subtitle Edit (version 3.5.1) was used for subtitling because it has a set of professional resources to create, adjust and synchronize subtitles besides allowing the incorporation of MT. The machine translation of the transcribed dialogues was carried out with a Google API Key application, generated in May 2020 and integrated into Subtitle Edit. The subtitles were post-edited by a translator/post-editor and revised by a second professional.

3.1.2 Material

The trailer selected for this study was *The Red Sea Diving Resort* (*Missão no Mar Vermelho* in Brazil), a 2019 production, directed

by Gideon Raff. According to IMDB (2019), the movie is inspired by true-life rescue missions, the story of a group of Mossad agents and Ethiopians who in the early 1980s used a deserted holiday resort in Sudan as a front to smuggle thousands of refugees to Israel. The undercover team carrying out this mission is led by Ari Kidron (Chris Evans) and Kabede Bimro (Michael Kenneth Williams). The movie trailer was selected from the catalog of Netflix Brazil in February 2020 and has 2 minutes and 21 seconds of duration.

It was selected according to the following criteria: 1. To be a movie trailer subtitled in the English-Brazilian Portuguese (EN/PT-BR) language pair; 2. To be available subtitled in the catalog of the streaming platform Netflix; and 3. The same trailer that was published on Netflix should be available for download on the Internet, but without subtitles so that the dialogues could be machine translated and post-edited into PT-BR.

3.1.3 Pilot Participants

Two pilot experiments were conducted: one with undergraduate students and another with translators. Students were recruited based on the following criteria: 1) they should be native speakers of Brazilian Portuguese; 2) have a preference for watching subtitled movies/series; 3) be undergraduate students of Languages at Federal University of Pernambuco and 4) have English level B1 (cf. Common European Framework of Reference). The translators, in turn, should have experience with translations in the EN/PT-BR language pair.

The experiment with students had two groups: 1) the control group that watched the trailer with the Netflix subtitles and 2) the experimental group that watched the trailer with post-edited subtitles. As the goal of this paper is to analyze the quality of post-edited subtitles, only the participants of the experimental group were considered for this analysis. Four students volunteered to participate in the experimental group, all female, aged between 19 and 23 years old, and undergraduate students of Languages at Federal University of Pernambuco.

The experiment with translators had 6 volunteers. Four of them were male and 2 were female aged between 18 and 50 years old with advanced or proficient English levels. All participants dedicated up to 10 hours a week to translation activity.

3.1.4 Experimental Design

Due to the Covid-19 pandemic, both experiments were conducted online. In the experiment with students, participants were asked to fill out a prospective questionnaire prepared in Google Forms. After that, they were directed through a link to watch the trailer with post-edited subtitles, and filled out a 5-point Likert-type scale on the same form. Then they answered a guided think-aloud protocol, which was recorded and transcribed for later tabulation and data analysis.

In the experiment with translators, initially the prospective questionnaire was filled in Google Forms. Then, through a link in the questionnaire, the volunteers were directed to watch the trailer with post-edited subtitles, followed by the completion of a 5-point Likert-type scale and the completion of open questions about the quality of the subtitles.

3.2 Data Analysis

The data analysis was based on qualitative and quantitative bias. To assess the quality of the post-edited subtitles, we followed the FAR model (cf. Pedersen, 2017). As explained in section 2.3, the FAR model is based on the analysis of errors to assess the quality of interlingual subtitling. In this model, errors are divided into 'minor', 'standard', and 'serious' according to the severity of the interference in the contract of illusion between viewers and subtitles, which can occur in three different areas: functional equivalence, acceptability, and readability.

Pedersen (2017, p. 217) states that the FAR model 'should be fed local norms, as presented in in-house guidelines'. In this sense, we

inform that until the date of this paper writing, Brazil did not have a national guideline for interlingual subtitling, causing Brazilian AVT companies to follow their own norms, which may vary from company to company. Thus, we adopted the norms indicated in the *Brazilian Portuguese Timed Text Style Guide* (Netflix) as parameters for the analysis of errors. We justify the choice of these guidelines because one of the criteria used to select the material was that it should be a trailer available on Netflix. Furthermore, Subtitle Edit was used to extract the data of the number of characters per second (CPS) and characters per line (CPL).

Both the guided think-aloud protocols and the open-ended questions were analyzed qualitatively in order to triangulate with the quantitative data from the Likert-type scale and the error scores (from FAR model).

All quantitative data received statistical treatment to analyze the differences and correlation between the variables analyzed in each group. The two pilot tests and the data obtained with FAR model application were run on SPSS software, and the cut-off point of significance adopted was $p \leq 0.05$. Following the methodology described in this section, we will present our analysis in section 4.

4. Analysis and Discussion

For the purposes of this paper, the analysis of interlingual subtitling quality assessment will focus on both data related to FAR model error scores as well as data collected with translators (quality assessment) and students (audience reception). The study aimed at analyzing the quality assessment of *The Red Sea Diving Resort* trailer subtitles, which were machine translated and then post-edited.

Figure 1 provides an overview of error scores identified in the category Functional Equivalence. The FAR model classifies errors into three categories: minor (0.5), standard (1.0) and serious (2.0).

Figure 1: Absolute Frequency of Error Scores in Functional Equivalence
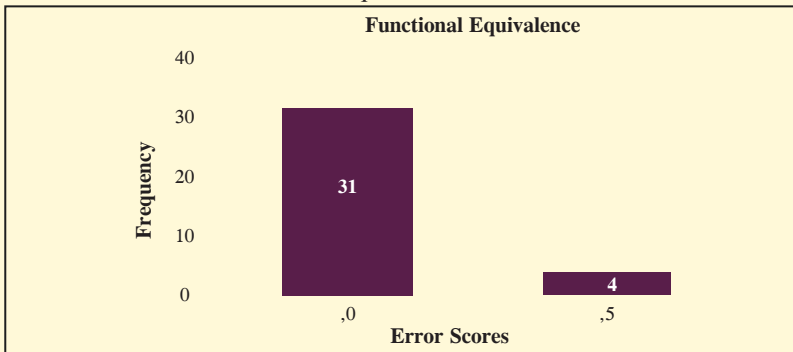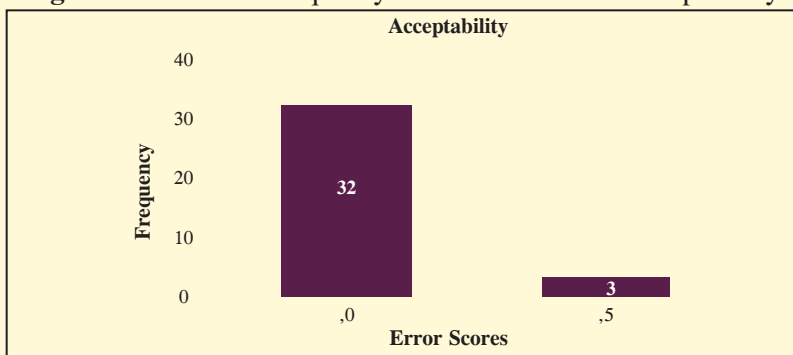

Source: Authors

As can be seen, the vast majority of subtitles had no functional errors (31 subtitles out of 35) and only four of them had minor errors. This result indicates that the meaning was well rendered in the translation and consequently no serious misinterpretations have occurred.

As stated before, Functional equivalence considers any type of error that would affect the meaning between the ST and the TT, and that includes dialects and other linguistic variations. The dialogues of *The Red Sea Diving Resort* trailer had little or non-culture-specific terms, so this result could present differences when analyzing the whole movie.

The analysis of error scores in the second category of FAR model, i.e. Acceptability, are presented in Figure 2. Acceptability errors are the errors that make the subtitles sound unnatural and can be subdivided into three types: grammar errors, spelling errors and errors of idiomaticity.

Figure 2: Absolute Frequency of Error Scores in Acceptability

Source: Authors

Similarly to Functional Equivalence, Figure 2 shows that most subtitles (91.4%) had no Acceptability errors. Minor errors were identified in only 8.6% of the subtitles. This result implies that the target text has conformed to target language norms.

On the other hand, the movie trailer subtitles were found to have high readability error scores, as can be seen in Figure 3. As previously explained, we adopted the *Brazilian Portuguese Timed Text Style Guide* (Netflix) to establish the error analysis criteria.

Figure 3: Absolute Frequency of Error Scores in Readability
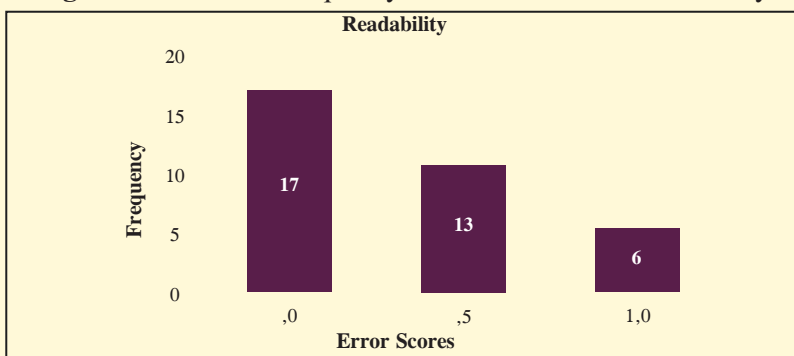
Source: Authors

Readability errors are related to technical norms or issues, such as segmentation, spotting, punctuation, reading speed, line length. The data shows that Readability errors occurred 52.8% of times. Most of them (36.1%) were minor errors and 16.7% were standard errors. This kind of error might affect the comprehension since Readability is related to how easy are the subtitles to the viewer to process them.

In a study conducted by Robert & Remael (2016) with 99 professional subtitlers, the participants admitted to follow the technical guidelines and "in their opinion, the most important parameters to affect quality were content, grammar, readability and contextual appropriateness" (Szarkowska, Díaz Cintas & Gerber-Morón, 2020, p. 4). The results of this study indicate that these types of errors play an important role in quality assessment.

Our analysis shows that Functional Equivalence and Acceptability had only minor errors whereas Readability had a higher score of severity, i.e., standard errors, which not only break the contract of illusion, but may also affect the subtitle in which the error is contained. Table 1 provides an overview of mean error score for each category.

Table 1: Mean Error Score for Each Quality Assessment Category of FAR Model

Functional Equivalence	Acceptability	Readability
.069 (SD = .17)	042 (SD = .14)	.347 (SD = .37)

Source: Authors

There was a statistically significant difference in error scores depending on the type of quality assessment category, $\chi2(2)$ = 22.706, p = 0.000. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p < 0.017. There was no significant difference between Functional equivalence and Acceptability error

scores ($Z = -0.707$, $p = 0.480$). However, there were statistically significant differences between error scores in Functional Equivalence *vs* Readability ($Z = -3.601$, $p = 0.000$) and between Readability *vs* Acceptability ($Z = -3.513$, $p = 0.000$).

These results regarding subtitles quality assessment provided by FAR model indicate that the post-edited subtitles have a good quality in terms of meaning and target language norms. Notwithstanding, the use of machine translation seems to have affected the conformity of subtitles to technical parameters, which impact on readability and consequently the way viewers process subtitles. In order to have a better understanding of the reception of the subtitles and whether viewers were affected or not by readability, we will now correlate FAR model results with audience reception as well as the translator's quality assessment of the subtitles.

We hypothesized that the lower the error score, the higher the satisfaction levels of the audience, which was measured by a 5-point Likert-type scale. A Spearman's rank-order correlation was run to determine the relationship between students' satisfaction levels and readability error scores, however the correlation was not significant ($rs = .000$, $p = 1.000$). This result may be due to the small sample size, so we also looked at the qualitative data, i.e., the guided think-aloud protocols. Among the questions, participants were asked to explain their rating on the Likert-type scale. 3 out of 4 participants (P01_E, P02_E and P03_E) mentioned no linguistic issue, but they pointed out technical parameters such as synchronization, subtitle font and color as having affected their appreciation of the movie trailer. P04_E, on the other hand, mentioned only a linguistic aspect, i.e., information omission, which is not an issue, but a very usual translation strategy employed in subtitling.

According to Pedersen (2017), one of the greatest weaknesses of FAR model is subjectivity when it comes to judging equivalence and idiomaticity errors and a degree of fuzziness when it comes to judging the severity of the errors. Taking that into consideration, we tested whether there was a difference in subtitles quality when

comparing FAR model scores and Likert-type scale rating provided by translators. A Mann-Whitney U test showed that there was a significant difference ($U = 0.000$, $p = 0.000$) between the quality results of FAR analysis (Mean Rank $= 18$) compared to the quality assessment provided by the group of translators (Mean Rank $= 38.5$). From this data, it can be concluded that translators' rating of quality was higher than the FAR scores.

This result should be interpreted with caution considering that FAR model is based on error analysis and does not reward excellent solutions (Pedersen, 2017), which could have been more prominent than the errors in the movie trailer. This explanation is supported by qualitative data from the open questions. When asked about any translation aspect that might have affected his/her comprehension and appreciation of the trailer, T04 verbalized: "If I am not mistaken, the subject (I) was omitted in one sentence, which caused some ambiguity and distracted me — and there was space to use more characters. Other than that, it flowed well"[4].

Additionally, when translators were asked to analyze the quality of subtitles, they watched the subtitled movie trailer only once whereas the analysts who applied the FAR model could watch the movie trailer as many times as necessary. Besides that, their judgment could be explained by the fact that good quality in translation is related to the "perception of a translation as most appropriate within the context in which it functions" (Bittner, 2011, p. 78).

5. Final Remarks

Based on the theoretical discussion about subtitling quality assessment, we can affirm that the product analysis process is a complex phenomenon, especially because of the multiple

---

[4] Translation from Brazilian Portuguese transcript of the open question: *"Se eu não me engano, houve uma omissão do sujeito em uma frase (eu) que deixou um pouco ambíguo e tirou meu foco – e havia espaço para usar mais caracteres. No mais, fluiu bem".*

connotations of quality. Furthermore, when discussing quality in an audiovisual product, we confirmed the eminent need to conduct further empirical-experimental research aimed at the final product. Consequently, the preliminary results of the study presented here and conducted by the GETRADTEC Group reached its primary objective: to analyze quality assessment of interlingual post-edited subtitling from an empirical standpoint.

Our initial analysis applied the FAR model, which made possible to assess functional equivalence, acceptability, and readability of subtitles. Our results presented here demonstrated that the post-edited subtitles had a good quality in terms of meaning and target language norms. However, the technical parameters had their quality affected by minor and standard errors, which could have not only broken the contract of illusion, but may also have affected the subtitle in which the error was contained.

As pointed out by Pedersen (2017), one of the greatest weaknesses of FAR model is its subjectivity when it comes to judging equivalence and idiomaticity errors and a degree of fuzziness when it comes to judging the severity of the errors. Thus, correlating the error scores with empirical data regarding quality assessment and audience reception has proven elucidating with regards to the concerns of avoiding subjectivity. Additionally, having collected data with the audience provided a clearer comprehension of whether viewers were affected by readability or not and showed that the quality assessment of translators was higher than the FAR scores, which might indicate that despite having some standard readability issues, the overall subtitling quality was not seriously affected.

Regarding the FAR model subjectivity, we noticed that penalizing errors as Minor (0.5), Standard (1.0) and Serious (2.0) might have contributed to its subjectivity. As a suggestion, we believe it might be helpful to have some intermediate error scores, such as 0.25, 0.75 and 1.5 to aid the categorization of errors on subtitles.

To conclude, the empirical results reported here must be considered in light of some limitations. The first is the small sample size, whose results may not be conclusive. Nonetheless,

it is noteworthy to say that this is an ongoing investigation and the same experiment is already being conducted with a larger sample. The second limitation concerns the heterogeneous sample of translators regarding professional experience, which may have affected their judgment of quality. This issue will be addressed in further experiments conducted by the GETRADTEC Group.

References

Abdelaal, Noureldin Mohamed. "Subtitling of Culture-Bound Terms: Strategies and Quality Assessment". *Heliyon*, 5(4), p. 1-27, 2019. DOI: https://doi.org/10.1016/j.heliyon.2019.e01411

Adão, Carlos Manuel Cunha de Jesus. *Tecnologias de Streaming em Contextos de Aprendizagem*. Master Thesis (Master in Information Systems). Engineering School, University of Minho, Guimarães, 2006. Repositorium UMinho. http://repositorium.sdum.uminho.pt/handle/1822/6400

Alfaro de Carvalho, Carolina. "Quality Standards or Censorship? Language Control Policies in Cable TV Subtitles in Brazil". *Meta: Journal des Traducteurs*, 57(2), p. 464-477, 2012. DOI: https://doi.org/10.7202/1013956ar

Athanasiadi, Rafaella. "Exploring the Potential of Machine Translation and Other Language Assistive Tools in Subtitling: A New Era?". In: Deckert, Mikołaj (Ed.). *Audiovisual Translation: Research and Use*. Bern: Peter Lang, 2017. p. 29-49.

Bittner, Hansjörg. "The Quality of Translation in Subtitling". *Trans-Kom*, 4(1), p. 76-87, 2011. Available at: http://www.trans-kom.eu/bd04nr01/trans-kom_04_01_04_Bittner_Quality.20110614.pdf. Accessed on: May 31, 2021.

Campos, Giovana Cordeiro & Azevedo, Thais de Assis. "Subtitling for Streaming Platforms: New Technologies, Old Issues". *Cadernos de Tradução*, 40(3), p. 222-243, 2020. DOI: https://doi.org/10.5007/2175-7968.2020v40n3p222

Dias, Murilo & Navarro, Rodrigo. "Is Netflix Dominating Brazil?". *International Journal of Business and Management Review*, 6(1), p. 19-32, 2018. Available at: http://www.eajournals.org/wp-content/uploads/Is-Netflix-Dominating-Brazil.pdf. Accessed on: Feb. 2, 2021.

Díaz Cintas, Jorge. "Subtitling". In: Millán, Carmen & Bartrina, Francesca (Eds.). *The Routledge Handbook of Translation Studies*. London & New York: Routledge, 2012. p. 273-287.

Doherty, Stephen; Gaspari, Federico; Groves, Declan; van Genabith, Josef; Specia, Lucia; Burchardt, Aljoscha; Lommel, Arle & Uszkoreit, Hans. "Mapping the Industry I: Findings on Translation Technologies and Quality Assessment". *QTLaunchPad*, p. 1-12, 2013. Available at: https://core.ac.uk/download/pdf/147606442.pdf. Accessed on: Feb. 17, 2021.

Doherty, Stephen. "Issues in Human and Automatic Translation Quality Assessment". In: Kenny, Dorothy (Ed.). *Human Issues in Translation Technology*. London & New York: Routledge, 2017. p. 131-148.

Gaspari, Federico; Almaghout, Hala & Doherty, Stephen. "A Survey of Machine Translation Competences: Insights for Translation Technology Educators and Practitioners". *Perspectives*, 23(3), p. 333-358, 2014. DOI: https://doi.org/10.1080/0907676X.2014.979842

Gottlieb, Henrik. "Subtitling". In: Baker, Mona (Ed.). *Routledge Encyclopedia of Translation Studies*. London & New York: Routledge, 2005. p. 244-248.

Gutt, Ernst-August. *Translation and Relevance: Cognition and Context*. London & New York: Routledge, 2014.

House, Juliane. "Translation Quality Assessment: Linguistic Description versus Social Evaluation". *Meta: Journal des Traducteurs*. 46(2), p. 243-257, 2001. DOI: https://doi.org/10.7202/003141ar

House, Juliane. "Quality of Translation". Baker, Mona (Ed.). *Routledge Encyclopedia of Translation Studies*. London & New York: Routledge, 2005. p. 197-200.

IMDB. *Missão no Mar Vermelho (2019)*. Available at: https://www.imdb.com/title/tt4995776/. Accessed on: June 2, 2021.

Netflix. *Brazilian Portuguese Timed Text Style Guide*. 2021. Available at: https://partnerhelp.netflixstudios.com/hc/en-us/articles/215600497-Brazilian-Portuguese-Timed-Text-Style-Guide. Accessed on: Apr. 18, 2021.

Nida, Eugene. *Towards a Science of Translating*. Leiden: E. J. Brill, 1964.

Nikolić, Kristijan. "Reception Studies in Audiovisual Translation – Interlingual Subtitling". In: Di Giovanni, Elena & Gambier, Yves. *Reception Studies and Audiovisual Translation*. Amsterdam & Philadelphia: John Benjamins, 2018. p. 179-197.

O'Brien, Sharon. "Towards a Dynamic Quality Evaluation Model for Translation". *The Journal of Specialized Translation*, 17, p. 55-77, 2012. Available at: https://www.jostrans.org/issue17/art_obrien.php. Accessed on: Feb. 7, 2021.

Pedersen, Jan. *Scandinavian Subtitles: A Comparative Study of Subtitling Norms in Sweden and Denmark with a Focus on Extralinguistic Cultural References*. Doctoral Thesis (Ph.D.). Faculty of Humanities, Department of English, Engelska Institutionen, Stockholm University, Stockholm, 2007.

Pedersen, Jan. "The FAR Model: Assessing Quality in Interlingual Subtitling". *The Journal of Specialized Translation*, 28, p. 210-229, 2017. Available at: https://www.jostrans.org/issue28/art_pedersen.php. Accessed on: June 2, 2021.

Pym, Anthony. "Quality". O'Hagan, Minako. *The Routledge Handbook of Translation and Technology*. London & New York: Routledge, 2020. p. 437-452.

Rabêlo, Melissa Silva Moreira; Garcia-Murillo, Martha & Couto, Carlos Agostinho Almeida de Macedo. "Public Broadcasting Services in the United States and Brazil: History, Funding and New Technologies". *Revista de Políticas Públicas*, 21(1), p. 469-494, 2017. DOI: http://dx.doi.org/10.18764/2178-2865.v21n1p469-494

Reiss, Katharina & Vermeer, Hans Josef. *Groundwork for a General Theory of Translation*. Translated by Christiane Nord. Tubingen: Niemeyer, 1984.

Robert, Isabelle & Remael, Aline. "Quality Control in the Subtitling Industry: An Exploratory Survey Study". *Meta: Journal des Traducteurs*, 61(3), p. 578-605, 2016. DOI: https://doi.org/10.7202/1039220ar

Romero-Fresco, Pablo & Martínez Pérez, Juan. "Accuracy Rate in Live Subtitling: The NER Model". Díaz Cintas, Jorge & Baños Piñero, Rocío (Eds.). *Audiovisual Translation in a Global Context – Mapping an Ever-Changing Landscape*. London: Palgrave Macmillan, 2015. p. 28-50.

Szarkowska, Agnieszka; Díaz Cintas, Jorge & Gerber-Morón, Olivia. "Quality is in the Eye of the Stakeholders: What do Professional Subtitlers and Viewers Think about Subtitling?". *Universal Access in the Information Society*, p. 1-15, 2020. DOI: https://doi.org/10.1007/s10209-020-00739-2

Toury, Gideon. *Descriptive Translation Studies and Beyond*. Amsterdam & Philadelphia: John Benjamins, 1995.

Arlene Koglin. E-mail: arlene.koglin@ufpe.br. https://orcid.org/0000-0003-0711-0811.

João Gabriel Pereira da Silveira. E-mail: jgsilveira96@gmail.com. https://orcid.org/0000-0003-3304-1768.

Morgana Aparecida de Matos. E-mail: morgana_matos@hotmail.com. https://orcid.org/0000-0001-5009-7638.

Vitória Tassara Costa Silva. E-mail: vitoriatassara26@gmail.com. https://orcid.org/0000-0002-1998-4316.

Willian Henrique Cândido Moura. E-mail: willianmoura.tradutor@gmail.com. https://orcid.org/0000-0002-2675-6880.