Quality Control in Audiovisual Translation

Raguž, Leonardo Ivan

UNIVERSITY OF RIJEKA

Faculty of Humanities and Social Sciences

UNDERGRADUATE THESIS

Quality Control in Audiovisual Translation

Submitted in partial fulfillment of the requirements for the B.A. in English Language and Literature and Philosophy at the University of Rijeka

Student: Leonardo Ivan Raguž

Supervisor: doc. dr. sc. Mirjana Borucinsky

Department of English Language and Literature

Academic Year 2021/2022

Abstract

Soon after the introduction of motion pictures, audiovisual media started gaining popularity rapidly and so the need for audiovisual translation began. Today, audiovisual translation is an invaluable part of media accessibility and many different modes have been tailored to complement various genres of audiovisual products. This paper describes the processes of several modes of audiovisual translation, focusing on subtitling and dubbing because of their popularity, but also covering the less popular modes such as video game localization, surtitling, voiceover and interlingual subtitling. As the title of this paper suggests, a major point of research is the concept of quality control in all of the mentioned modes of translation. More specifically, how quality control is performed, who performs it and what guidelines and standards should be followed in order to produce a high-quality audiovisual translation.

KEY WORDS: audiovisual translation, media accessibility, modes of audiovisual translation, quality control, subtitling, dubbing, video game localization, surtitling, voiceover, interlingual respeaking

Table of Figures

List of Abbreviations

AVT – audiovisual translation

IRSP – interlingual respeaking

Table of Contents

1. Introduction

Audiovisual (AV) products, as the term suggests, are meant to be watched and listened to at the same time. This means that, unlike typical translations of literary texts or academic papers which transpose only the written word and its meaning from one language to another, audiovisual translations need to take into consideration the interaction between word, image, sound and other numerous components of audiovisual products (Taylor, 2013, p. 99-100).

Audiovisuals are available in many forms, some of which are movies, TV shows, video games, theatre, etc. Accompanying these AV products are also numerous manners in which they can be translated, such as subtitling, dubbing, voiceover, surtitling (subtitling for opera and theatrical plays), video game localization and respeaking. While there are more approaches to AV translation than the ones listed above, these are the modes which will be elaborated in this paper, with a major focus on subtitling and dubbing.

While subtitles are usually made by only one trained translator, a technician and, potentially, a quality controller and do not require many technological resources, dubbing demands a whole team consisting of one or more translators, voice actors, dubbing assistants, a dubbing director etc. as well as studio equipment which results in a much higher production cost than that of subtitling. This is one of the reasons why subtitling is generally a more accepted and more well-received practice among viewing audiences. Which is the more common method of language transfer varies greatly depending on which country we look at, for example, in France, Germany, Italy, Spain and Hungary dubbing is the preferred, or more accurately, more common method of AV translation. On the other hand, countries like Belgium, the Netherlands, Scandinavian countries and of course Croatia mostly use subtitling as the preferred translation method. Both of these methods face a number of issues (such as censorship, textual reduction, isochrony[1], lip synching) that translators and various technicians must do their best to overcome in order to produce a quality translation (Chiaro, 2009, p. 151). Such issues will be further elaborated in this paper.

Quality control, although still not thoroughly researched, is an extremely important part of any formal act of audiovisual translation. It can greatly affect the final product of audiovisual translation as well as improve the viewer's experience while consuming audiovisual products.

[1] Isochrony is the postulated rhythmic division of time into equal portions by a language. (https://www.definitions.net/definition/ISOCHRONY)

The term ‘quality’ is used very often in everyday life, almost everything is evaluated and compared according to its quality, even objects as simple as our toothbrushes or even chairs. Things are no different when it comes to AVT, attention to detail is key in producing a satisfactory translation.

For example, while producing subtitles for a movie, other than correctly transferring meaning of words and sentences from one language to another, the timing of each individual subtitle appearing on screen and disappearing from it is of utmost importance (Nikolić, 2021, p. 67). On the other hand, in dubbing, the synchronization of lip movement and the actual recorded speech must be incredibly accurate, the loudness and intonation of characters’ speech has to correspond perfectly to their facial expressions as well as their gestures. These requirements make AVT very complex which is why quality control is of such importance. Many experts have already attempted and are still attempting to improve upon the concept of quality control in AVT. However, that is quite a difficult task to accomplish since a system for measuring quality would need to eliminate all subjectivity of the one performing quality control to be completely accurate. Such a system does not yet exist but with the rapid advancement of technology it might become a realistic possibility.

The main complexity of AVT in general is its multimodality, which has “become the central framework for research on audiovisual texts, i.e. texts which create meaning through the use of multiple semiotic modes, such as films.” Those semiotic modes are; the aural-verbal mode (dialogues and lyrics); the aural, non-verbal mode (music and sound effects); the visual-verbal mode (various types of text on screen); and the visual, non-verbal mode (images). Multimodality attempts to blend these semiotic modes into a single whole by examining their exact functions (Remael & Reviers, 2019, p. 260).

The aim of this paper is to detail the processes and differences of various modes of AVT and to show how quality control can be performed to ensure an easier and more enjoyable viewing experience. Focus will be put on subtitling and dubbing which are arguably the two of the most researched and well-known modes of AVT.

2. Modes of Audiovisual Translation

As mentioned earlier, audiovisual translation comes in many different modes. This chapter will focus on brief elaborations of some of the modes of AVT, excluding subtitling and dubbing, to which separate chapters are devoted. In particular, a closer look will be taken into video game localization, surtitling, voiceover and interlingual respeaking.

2.1. Video Game Localization

Video games are computer-based entertainment software involving a single player or any number of players in a virtual environment. In current times, a very large number of video games from many genres is available for purchase online and in physical stores alike. This industry has already overtaken TV and cinema in terms of earnings, and not by a small margin. One of the reasons for that might just be the attention that game production companies give to quality control during localization. Unlike in most of the other modes of translation, in video game industry translation is an integral part of production and not just something that is done after production has ended, it is not an "unrelated and unimportant appendage", translators are involved in every stage of the project (Chiaro, 2009, p. 153).

Localization is a term used to describe the technical, cultural and linguistic adaptations made to the original software. Completing localization requires much more than text-to-text translation. Creativity is a valued virtue in a team of localizers, since video games combine numerous storytelling techniques which are used in both literature and cinema with player interaction cues during gameplay. Players themselves influence the way the story unfolds which is why traditional text translation is not enough to make a game playable. That is, grammatical correctness on its own does not provide a satisfactory playing experience for the target audience. Text translation is only a part of "a complex, polysemiotic[2] whole. In other words, you translate a text, you localize a video game." (Bernal-Merino, 2020, p. 299)

Passive media such as works of literature, movies, TV, etc. may contain up to five semiotic layers: spoken language, written text, image, music and sound. Communication between the reader/viewer and the consumed product is unidirectional. On the other hand,

[2] Polysemiotic – containing multiple semiotic (semiology – study of signs and sign-using behaviour) layers (https://www.definitions.net/definition/semiotic)

interactive entertainment media is necessarily bidirectional. That is, players' decisions influence the video game on the spot in such a way that no two playthroughs are ever the same. So, other than the five semiotic layers already mentioned, three new layers are introduced in interactive media, these are touch, proprioception (sense of body movement and position) and equilibrioception (ability to experience and maintain physical balance). A great example for this is the flight simulator, in which the cabin rotates and the controller vibrates according to what the player does. This bidirectional communication between game and player serves as a guide for localizers to improve playing experience for each specific target locale.

If a team successfully completes localization and all semiotic and pragmatic layers of the product are coherent within the game world, it can be said that "unicity" has been achieved within the game. In other words, the game provides a unique satisfactory experience for players. But, since the game has to be adapted for players all over the world, unicity can easily be shattered. It is not a simple task to retain the same level of coherency in every version of the game, hence errors in localization are not uncommon. The most common mistakes can be categorized in four groups: "language environment (spelling mistakes, lexical and syntactical calques, taboo language, poor translation style), sound environment (flat performance by actors, inappropriate music, asynchronous recordings), visual environment (culturally offensive graphics, inadequate user interface font size and color, unreadable subtitles and captions, confusing icons, inappropriate body gestures), pragmatic environment (unresponsive controls, mismatch between triggers and reactions, culturally inadequate gameplay demands)." (Bernal-Merino, 2020, p. 300-3) When working on such a creative medium, collaboration between translators and developers is extremely important, as well as quality control.

In the video game industry, quality assurance is a necessary process. Final proofreading, identifying and registering certain errors (grammar, spelling, text-image synchronization, text bleeding, etc.) is done by linguistic play testers. During linguistic quality assurance numerous different bugs might be found, such as: legal bugs (license agreement, end-user agreement), issues in coding, incorrect translation, audio issues (cut-offs, incorrect timing), out of context translation, wrong text display and so on. Although perfection is desirable, often it is not a priority, pragmatism is usually the goal in centralization. According to this, game bugs are divided into four categories depending on how a certain bug influences the game. These categories are A- unable to sell the game, B- bug affects playability but game is playable, C- does not affect playability and D- barely noticeable. Taking into consideration all the stages of linguistic quality assurance, workload has become so high that certain firms (e.g., PoletoWin)

work solely on this part of game localization. More precisely, such firms develop software which helps to identify and track bugs to allow for simultaneous testing on different language versions of the video game, so they can all be finalized and ready for release at approximately the same time (Bernal-Merino, 2020, p. 305-6). The focus and importance that is put on quality control in localization is evident, it is an invaluable part of video game production.

2.2. Surtitling

In the Oxford Advanced Learner's Dictionary the word 'surtitles' is defined as "words that appear on a screen above or next to the stage to show or translate into a different language what is being sung in an opera, or spoken in a play in the theatre". As the definition states, surtitles provide a translation of a live opera show or a theatre play from a source language in which it is performed to a target language the audience can understand. Such a translation is usually projected on a screen above or beside the stage. It carries an evident resemblance to cinema subtitling because of the way in which it is presented to the audience, but the two practices also share similar spatial and temporal limitations, as well as translatory constraints (the translated text must be condensed and fragmented) (Carrillo Darancet, 2020, p. 174). Because of these constraints every detail must be considered in order to find the best way of transferring the intended meaning with minimum disturbance of the original text.

Surtitles first appeared in the 1980s in Canada with the Canadian Opera Company producing the first ever surtitled opera in 1983, while the New York City Opera was the first to implement it systematically. Very soon after that, opera companies and theatres around the world started adopting this practice. With the increasing availability of subtitles and quick technological advancements, surtitling is gradually becoming a default as well (Secară, 2019, 130-131).

Although they share many similarities, unlike subtitling, surtitling requires the translator to manually change each caption which appears on the screen with the help of a suitable computer program while following the play. This makes the translator an important part of the live performance, which puts them under harsh scrutiny in the event of an error. Surtitling may be prone to judgement by authors and audiences because of the necessary text reduction which omits certain elements of the original text so that the captions can more easily fit on the screen;

this reduction means that some parts, however subtle, will be lost in translation. In addition to a few possible mistakes that potentially occur in the surtitling process (typing mistakes, timing errors or lack of synchronicity between actors' speech and the surtitles), these are the main reasons why this practice can face harsh criticisms (Carrillo Darancet, 2020, p. 174-76). It is quite difficult for a translator to meet all of the demands of the production company and the audience.

Surtitling may be divided into three stages, the actual translation of text, adaptation and projection. All of these stages are heavily reliant on technology, namely the computer software which allows the translator to set the number of characters as well as adjust the time allotted for each title, and the projection hardware that the production has decided on. Because of this dependence, rapid technological advances have allowed for many different types of surtitle projections that affect readability of the titles. Usually, surtitles are projected onto a screen made of cloth or a special fabric, but with time and increasing budget new techniques have been developed involving LED panels, screens on the back of seats, projections on mobile phones, glasses which project surtitles to individual viewers, etc. All of these techniques have serious disadvantages, such as distracting the viewers from the live performance, comfort while wearing or carrying viewing devices and various technical issues (weak Wi-Fi signal on the glasses or phones, inability to set up the device) (Carrillo Darancet, 2020, p. 180-81). This field still requires a lot of research, but modern technology has already brought many improvements.

The method of surtitling is also very important in making operas and plays accessible to those who are deaf, deafened or hard-of-hearing. Some may think the number of people with such disabilities is not very high, but it must be taken into consideration that hearing loss is a quite common occurrence in older age (statistics show that over 40% of people over the age of fifty experience hearing loss to some degree). It is difficult for many of these people to fully enjoy plays without some form of support, so titles are invaluable to them (Secara, 2019, p. 131).

2.3. Voiceover

Voiceover is a mode of audiovisual translation which has not yet been researched as thoroughly as subbing and dubbing but is still commonly utilized in the market for AVT (Chiaro, 2009, p. 152). It has sometimes been called the "ugly duckling" or the "orphan child" of audiovisual translation because of its limitations and audience reception when used in fiction such as films and TV shows. In some countries, voiceover is a common practice because of its low cost, although better alternatives are often available (at least when considering fictional content). There are a few different manners in which voiceover can be produced, and since it is used in both fictional and non-fictional genres of audiovisual products, a lot of research still needs to be carried out in this area, which has begun to increase during recent years. The next few paragraphs will briefly elaborate the varieties, main characteristics and some issues of voiceover.

The word 'voiceover' seems to have two slightly differing definitions depending on which field of research is being covered. On the one hand, in translation studies, it is defined as a method in which a voice is heard uttering a translation in a target language simultaneously over the source language voice. It is also specified that in most cases original audio volume is decreased (although still hearable) so that the translation is more easily heard and understood. Although the volume is decreased eventually, it is common to let the audience hear the beginning and ending of the original speech at full volume to increase the feeling of authenticity. This means that the translated speech begins a short time after original speech begins, and translation ends a short time before the original speech ends.

On the other hand, in film studies voiceover is defined somewhat differently, as a voice of a narrator which is offscreen or a voice which is heard but does not belong to any character talking in the scene. Film studies indicate that in documentaries or newsreels the voiceover will usually be that of a commentator who gives the viewers a third-person overview, an expert explanation of what exactly is unfolding on the screen (this is sometimes called voice-of-God narration). In films portraying fiction, voiceovers may also be used to give the audience an insight into a character's thoughts and mental states, as well as for directors or other experts to provide comments during films in certain DVD releases (Matamala, 2019, p. 64-65).

Despite the lack of research, some internal categorizations of voiceover have already been provided. For example, difference has been identified between voiceover for production and voiceover for postproduction. The difference being whether translators are given a

complete, already airable program (voiceover for postproduction), or are they given a not-yet-edited product whose excerpts they have to translate (voiceover for production). Another basis for voiceover categorization is the number of voices (or voice actors) present in the translation. This categorization features single-voice and multiple-voice voiceovers. An example of single-voice voiceovers can be observed in Poland where all characters of certain television shows are voiced by a single person. The latter variety can be found in Spain where certain documentaries are translated in such a way that numerous voices are incorporated to replace original speakers (Matamala, 2019, p. 67). Of course, multiple-voice voiceovers are usually more effective and better received.

In the past quite a few countries used voiceover systematically as a means of translating fictional genres, while nowadays it is less common but still sometimes used in certain regions, although majority of viewers prefer other modes of audiovisual translation. For example, it is still often used in a number of eastern European countries. Poland is one of the strongholds of voiceover, where voiceover is used for both fictional and non-fictional programs. According to a BBC survey in the early 2000's voiceover was a preferred mode of audiovisual translation by the viewers there (52% of viewers preferred voiceover over other modes), however that has changed as more recent observations show otherwise. Television broadcasting in the Ukraine also commonly utilizes the voiceover technique, while in Russia, it is heavily used in fictional and non-fictional genres on DVD's, cinema and TV (Matamala, 2019, p. 70-71). In Croatia, which is mostly a subtitling country, voiceover can be found in combination with subtitling in non-fiction (documentaries, news broadcasts).

As is evident from its definition, the main characteristic of voiceover is the translating voice being heard over the translated original audio. But, this is not the only important observable characteristic, as various synchronies are included in voiceover translation, such as: voiceover isochrony (the influence the duration of original speech has on the translating speech), kinetic synchrony (synchronization of translation with body language and gestures of characters) and action synchrony (synchronization of translation with visuals on screen). When these synchronies are successfully retained, the voiceover quality substantially increases. It is also important to note that unlike in dubbing, voiceover does not include lip synchronization in the translation, the audience is always simultaneously presented with both the original and translated audio tracks. Although it is much cheaper and faster than dubbing, voiceover faces many issues, including textual reduction, omitting and rephrasing information, possible lack of comprehension because of overlapping audio, socio-cultural adaptation during translation etc.

(Matamala, 2019, p. 67-69) Voiceover will always retain an important role in translating documentaries, interviews and newsreels, but it is evident why in fictional genres its popularity has decreased heavily. The presence of two simultaneous speeches hinders comprehension and the lack of lip synchronization can interfere with the atmosphere an audiovisual product is supposed to produce. There is a reason why it has sometimes been referred to as the "ugly duckling" of audiovisual translation.

2.4. Interlingual Respeaking (IRSP, Live Subtitling)

Respeaking is a method which utilizes speech recognition software in order to repeat or paraphrase that which is heard from an audiovisual text while also saying aloud every necessary punctuation mark and adding special features if needed (e.g., colors to discern between different speakers) (Dawson, 2019, p. 37). The respeaker utterances are then turned into onscreen text and are cued as live subtitles. The technique of intralingual respeaking has already been well established in translation studies, whereas interlingual respeaking (IRSP) has not yet been widely implemented as a method of translation. The first use of intralingual respeaking happened in 2001 in the UK, when it was used during the airing of BBC World Snooker Championships. Nowadays it is used frequently, primarily to make audiovisual content more accessible to those who are deaf or hard-of-hearing. IRSP mimics the same process as intralingual respeaking but adds another layer of complexity to an already difficult practice, live translation from the source language to a target language. This method is still not widely practiced and further research is necessary before live subtitles can be produced with consistently better quality.

IRSP is a new mode of audiovisual translation and as such is still in its early stages of development. A few recent studies have attempted to research in detail the intricacies of this method by recruiting translators with previous experience in translation, interpreting and intralingual respeaking. Two of such studies are T*he Interlingual Live Subtitling for Access* (ILSA) project (2017-2020) and the *SMART study* (2018). The goal of the former was to develop and validate the first training course for IRSP, while the latter compared the performances of trainees in IRSP (Dawson, 2019, p. 37). Interest for improvement in this discipline is increasing, and rapid technological advancements of speech recognition software are an important factor in making that improvement a reality.

The existing research on intralingual respeaking has helped immensely in developing parts of training courses and quality standards for IRSP. The area of research where this is evident is the specific skills that must be acquired or improved in order to effectively respeak an audiovisual text. The skills needed are: Research-mining, cultural knowledge, dictation (and punctuation), multitasking, live translation, language, source language comprehension, target language expression, error correction, editing, short-term memory, critical analysis and reflection (Dawson, 2020, p. 219). These skills mostly overlap in both intralingual and interlingual respeaking. The overlapping skills should not be ignored because they might need to be relearned or adapted according to context of the audiovisual text, as proficiency in certain skills may prove to be insufficient when changing from intralingual to interlingual respeaking. The cause of this is the layer of complexity which is added during language transfer process.

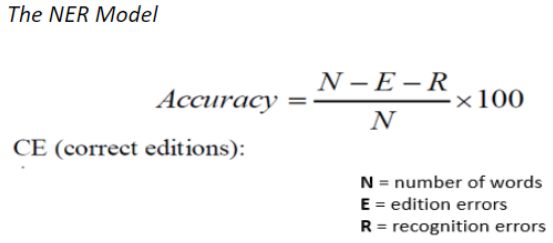
Another important factor of all types of respeaking is the software and hardware with which is done. Practical experience is extremely important for respeakers to get acquainted with all of the features of the software and hardware, so they can make necessary changes to the text and correct potential errors as quickly as possible (Dawson, 2020, p. 221-22). The ability to interact with technology is one of the most important aspects of live subtitling, since speech recognition software is not yet perfect and can make a mistake during recognition which has to be corrected, but also during the respeaking of dialogues it is necessary to provide distinction between the utterances of different speakers, this makes the dialogue much easier to understand.

One of the central topics in both intralingual and interlingual respeaking is quality control in live subtitling. The quality of live subtitles is affected by key factors which include subtitle latency, speed and accuracy. Live subtitles generally lag behind the original speech (subtitle latency), the duration of the delay is dependent on the original speaker's speech rate, the respeaker's speech rate and the reading rate of the viewers. If the respeaker is translating the original speaker's speech verbatim, it means he has to they have to literally utter more words than the person whose speech they are respeaking (because of punctuation enunciation and error correction). Therefore, original speeches are usually paraphrased to reduce the total delay. For example, if the original speech's speed is around 180 words per minute, the respeaker will usually fall behind by a margin of 0-20 words, and this delay can only get worse with speed increase. So, some editing is necessary, especially if the original speaker's speed is very high. In addition to the human-related delay, the speech recognition software also has its own delay as it takes a small amount of time to process information (Davitti & Sandrelli, 2020, p. 110).

Decreasing latency is always an area of concern since live subtitles should correctly correspond to what is shown on the screen.

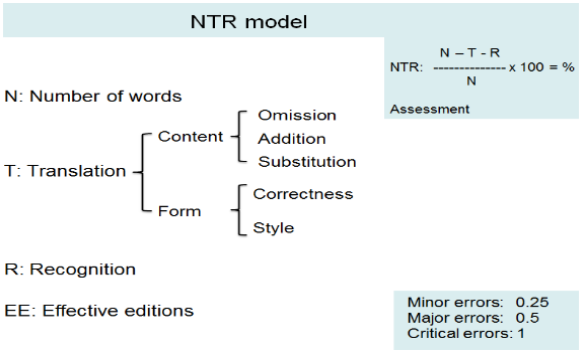
Two similar formulas have been developed to measure accuracy in live subtitling: the NER model used for intralingual respeaking (Romero-Fresco & Martinez, 2015, p. 32) and the NTR model used for IRSP (Romero-Fresco & Pöchhacker, 2017, p. 159). The former model is designed to differentiate between the respeaker's editing errors and the software-based errors of recognition. The errors are then categorized according to their severity, which can be minor, standard or serious errors. Minor errors do not seriously hinder comprehension of the text and, thus make the text comprehensible; standard errors are those that induce confusion or omit a piece of information; and serious errors provide the viewer with incorrect and misleading information. Those editions which do not contain loss of information are considered to be correct editions and are not included in scoring the accuracy but can be used in post-analysis to register the respeaker strengths. If the accuracy score of a respeaking task is over 98%, it is considered as an acceptable result. The NER formula in Figure 1 is used to calculate the accuracy rate of the text; the N stands for 'number of words', the E stands for 'editing errors' and the R stands for 'recognition errors'. The errors are determined by the person measuring the accuracy.

Figure 1 The NER Model Formula



Source: Romero-Fresco and Martinez (2015, 32)

Figure 2 The NTR Model Formula



Source: Romero-Fresco and Pöchhacker (2017, p. 159)

Almost the same criteria apply to the NTR model as well, which is based on the NER model. The main difference is that the NTR model differentiates between recognition errors and translation errors (unlike the NER, which considers edition errors). Translation errors are

those content-related (omissions, additions and substitutions) and form-related (grammar and style errors). Just as in the previously described NTR model, the errors are categorized by their severity: minor errors which can be easily understood; major errors which contain loss of information or induce confusion; and critical information which provide the audience with incorrect information (Davitti & Sandrelli, 2020, p. 110-12). The NTR formula in Figure 2 is used to calculate the accuracy rate of the live translation; here N is also a symbol for 'number of words' and R for 'recognition errors', but the T stands for 'translation errors'.

These techniques of quality evaluation can be very helpful in training respeakers as well as improving overall quality of live subtitles by analyzing results and developing new standards in respeaking. But the methods are far from perfect as they are based on the raters' evaluation and do not feature assessment of the actual viewers via a reception study. Also, sometimes it is not so easy to determine whether something is a correctly used strategy or an error and vice versa, meaning that error observation is partially subjective. Therefore, this type of quality control, while helpful, may not be completely precise and further research and improvement are required.

3. Subtitling

This chapter will focus on explaining the process of producing subtitles and how and why quality control is carried out in the subtitling industry with the help of some data from two research articles published in 2016 (Robert, Remael) and 2021 (Nikolić). Alongside dubbing, subtitling (also referred to as subbing) is one of the two most common modes of audiovisual translation, and consequently, the most researched ones.

Subbing is a method of audiovisual translation that "consists of rendering in writing, usually at the bottom of the screen, the translation into a target language of the original dialogue exchanges uttered by different speakers, as well as all other verbal information that appears written on-screen (letters, banners, inserts) or is transmitted aurally in the soundtrack (song lyrics, voices off)." Unlike dubbing, subtitling retains the original text, both aural and visual, while also giving the viewers another layer of information. Therefore, all subtitled audiovisual products include the following three layers of information: 1. the original spoken and written word, 2. the source image and 3. the text of the subtitles. Because the material subtitlers work with incorporates these multiple semiotic layers which are both aural and visual, they need to come up with solutions which will achieve balance between these components. To do so, the speed at which viewers read the subtitles while simultaneously looking at the images must be taken into consideration; such obstacles reveal the constraining nature of subtitling (Diaz-Cintas, 2020, p. 150-51). Therefore, different guidelines have been introduced to simplify the subtitling process, this paper will take a look at the guidelines proposed for subbing in Croatia.

Many different norms for assessing subtitling quality have been proposed in various countries and corporations, but no one strict universal norm has been systematically implemented. Study and development of methods for quality control in translation did not start in recent years. In as early as 1977, Julianne House published an article and later a book introducing a model for evaluation of translation quality based on her doctoral dissertation. Since then, many other academic works have been published on this topic, including some other models for quality assessment, but such models cannot successfully be utilized in quality control of subtitles because of the added spatial and temporal obstacles that they face. That is why audiovisual translation experts have designed methods specifically for quality control of subtitles (Nikolić, 2021, p. 68-69). The NTR and NER models that have already been mentioned are widely used in live subtitling. Some problems still exist within these and other models as they can be very time consuming and not fitting for long audiovisual content. This is why a

survey was conducted by Nikolić in 2021 among professionals involved in different stages of subtitle production, which sheds some light on the current state of quality control in the subtitling industry. One of the parts of this chapter will look at the results and impact of this research.

3.1. The Process of Subtitling and its Quality Guidelines

Presented in a brief and simplified way, the subtitling production process looks something like this; it starts with a broadcasting or a streaming and video-on-demand (SVOD) company putting in a request for a certain audiovisual product (movie, TV show, documentary, reality show, etc.) to be subtitled; that request then generally goes to a language service provider (LSP); the LSP hires professional subtitlers to produce subtitles for the AV product (unless, as sometimes may happen, the broadcaster or SVOD directly hires subtitlers, depending on their business model). After the subtitle text is written, the stage of quality control ensues. Quality control (QC) is in most cases done by very experienced subtitlers who proofread the subtitles and check their technical accuracy in accordance with company guidelines. Of course, spelling, grammatical correctness, register and translation accuracy are also checked by quality controllers (Nikolić, 2021, p. 67). At times, thorough quality control is not carried out, usually for money saving purposes.

Through the website of Audiovisual Translators Europe (AVTE) numerous national AV translation associations' websites can be reached (German, Italian, Portuguese, Spanish, UK, Croatian, Slovenian, etc.). The main purpose of those websites is to make the recommended subtitling quality guidelines accessible to translators and the public. Those quality guidelines suggest how subtitles should look, function and in what cases certain adaptations should be used (Nikolić, 2021, p. 70).

'Guidelines for Quality Subtitling in Croatia' (avteurope.eu), as the name suggests, is a document available on AVTE's website which contains the recommended rules that are used during subtitle production. The guidelines are split into two parts; the first one pertaining to language aspects and the second pertaining to technical aspects of subtitling.

The first section, or the language section, begins by explaining the most basic of all elements, namely grammar and spelling, which must be correct unless the audiovisual content allows for departure from the standard Croatian language. The next item explained is the

translation of idioms; the idiomatic part of AV content should be matched in style during subtitling and adapted to the target audience (this includes figures of speech, slang, curses, etc.). During a dialogue between two or more characters, differentiating between speakers is necessary, so a dash is put before the utterance of not the first, but the second and all subsequent speakers involved in the dialogue. Longer sentences should not go over into several subtitles but should rather be cut into several shorter sentences if possible (sometimes the structure of the audiovisual content does not allow this). In some cases, subtitles can also be italicized, this should be minimized so as not to further confuse viewers which are already processing visual and aural information. Italics are used for lyrics of songs, names of hotels, cars, businesses, books; emphasized words, quotes and reported speech, narrator's comments and so on.

Good subtitles do not only transfer the meaning of what is spoken, they also need to mimic the style and tone of speech, all the while being unobtrusive and enabling the viewer to follow the plot with ease. Audiovisual content which is completely irrelevant for the story should not be translated. For some audiovisual products, correct terminology is quite important; for example, technical terms in certain TV shows should be translated identically through all episodes. When multiple foreign languages are spoken in a single scene, only when the characters in the scene do not understand the other foreign language speaker are his subtitles put in brackets (in other cases, brackets aren't necessary). This concludes the language aspects of the Croatian quality subtitling guidelines.

The technical aspects of subtitling are defined by the visual appearance of the subtitles, their spatial and temporal limitations. Arguably the most important aspect, although simple, is the readability of subtitles. In Croatia, and many other countries, subtitles are written on a transparent gray background with white letters to make them as easily readable as possible. They are located in the middle of the bottom part of the display, with the exception of descriptive subtitles which are placed right above the usual subtitles. The text of the subtitles should be displayed for a long enough time for the audience to read it; standard duration is two and a half seconds for a single line and 4-5 seconds for a two-line subtitle, this can be adjusted depending on text difficulty and target audience. Time coding is another vital part of subtitling, as subtitles must appear simultaneously with the speech they provide the translation for, but also should not cross into following scenes (this can be tolerated for extremely short scenes). In addition, it is standard to leave a pause of at least three frames between two subtitles for the viewer to notice the change (avteurope.eu).

All these suggested directives are designed to make a subtitler's job simpler. Although that goal is partially fulfilled, subtitlers are still left with various difficult tasks. Possibly the most difficult and vital task of a subtitler is condensing the dialogue, that is, adapting the translation in such a way as to decrease the amount of text the viewer must read. Two types of such text reductions exist; partial (paraphrases or condensation) and total (deletions). Partial reductions are methods of successfully transferring the same amount of spoken information in a shorter text by phrasing said information in a different manner. On the other hand, deletion involves completely omitting unnecessary words or phrases from the translation, such as markers of interaction (tag questions, subject ellipsis, interactional terms), 'gestural' language (exclamations, affirmations/negations, fillers), repetitions, background chatter, some deictic expressions and proper names (de Linde, 1995, 15-16). For example, certain words and phrases are omitted to save space; such as "listen to me", "ugh" or "ew" (expressions of disgust), "… isn't it?" (endings of tag questions). Both types of reduction are commonly used in subtitling and effectively help the subtitler to produce a quality translation.

3.2. Professional Perspectives on Quality Control in Subtitling

Quality is one of the main topics in subtitling and can be approached from various angles. Since many parties are involved in the production and utilization of subtitles, their understanding and quality demands can vary. Broadcasters and SVOD companies might at times put more emphasis on the quality of technical aspects in subtitles, while subtitlers themselves are generally more focused on linguistic aspects (grammar, style, correct and condensed translation). Therefore, quality control is not always carried out with the same attention to detail on all the aspects.

Quality control can be affected even before the actual process begins by providing the subtitler with appropriate technical equipment for the job. During the translation process, quality is managed by performing revision ('checking') on the subtitles by the subtitlers themselves; and after the translation process by revision which is performed by another subtitler (Robert & Remael, 2016, p. 579-80).

Surveys have been conducted among professionals to determine how often and with how much attention to detail these types of quality control are performed. Robert and Remael have published an article in 2016 containing the results of an online questionnaire regarding

quality control which was filled out by 99 subtitlers. Although the completed survey can only offer a glance into the researched topic, given the scope and the quickly evolving nature of audiovisual translation, it provides helpful insights which can be used for improvement of quality standards in the field.

The analysis of the results showed that as much as 45% of the surveyed subtitlers revised their own work in one or multiple steps, with the revision process varying greatly. However, when it comes to revision of others' subtitling work, 39% of the respondents reported that they had never revised others' subtitles; this statistic does not meet the suggested standard in the subtitling field. This is surprising because many clients do offer revision instructions for both technical and linguistic aspects of subtitles (even though linguistic aspects are generally less detailed) as well as technical equipment for such tasks. It can be concluded that the clients do not share the same understanding of quality as the actual translators, nor the need for extensive linguistic quality control, perhaps because of certain economic crises and the increasing competition in the field (Robert & Remael, 2016, p. 600-01).

Another more recent article was published in 2021 by Nikolić on the same topic. The questionnaire reported on in this article did not focus solely on subtitlers but on all stakeholders involved in the process of subtitling (including quality controllers, employees of language service providers, broadcasters and proofreaders). Each group of stakeholders answered different questions regarding quality control in the subtitling industry. Two main hypotheses were to be checked by the questionnaire: A – different stakeholders hold different opinions on the meaning of 'quality' of subtitles, and B – different stakeholders have different opinions on who holds the most responsibility for quality of subtitles.

Unlike the questionnaire conducted by Robert and Remael, analysis of results from this more recent survey has yielded very different results when it comes to its hypothesis. Nikolić came to the conclusion that "…stakeholders do not have significantly different opinions about what constitutes quality subtitles." (2021) The answers obtained imply that all stakeholders put similar emphasis on both technical and linguistic aspects of subtitles, with readability being mentioned quite often as arguably the most important parameter of subtitles.

When it comes to the second hypothesis, which stakeholder is the one most responsible for the quality of subtitles, answers unsurprisingly slightly vary since they look at quality control from differing points of view. Although the majority of respondents list the subtitlers as most responsible for the quality of subtitles, certain broadcasters and language service

provider employees believe that quality controllers carry more responsibility (Nikolić, 2021, p. 85).

Both articles mention that not all language service providers or broadcasters put the same emphasis on quality. Eighteen subtitlers said in the survey that quality control processes (whether they exist or not) depend on the client. Which means that in some cases, a subtitler's own revision is the only form of quality control.

4. Dubbing

Dubbing is a popular mode of audiovisual translation defined as a procedure which includes replacing the original speech of an audiovisual product with a voice track translated into the target language, while trying to mimic lip movements, phrasing and time code of said original speech. The main goal of dubbing is to make it appear as if the translated speech is being uttered by original voice actors in order to make foreign audiovisual products enjoyable and understandable for the target audience. It can be divided into four stages. First, the screenplay is translated; second, adjustments are made so the translation fits in with the original lip movements and so it sounds natural when recorded in the target language; third, voices are recorded according to the newly translated screenplay; and finally, the translated audio track is edited into the original recording (Chiaro, 2009, p. 144-45). Dubbing is quite a unique manner of translation in the sense that delivering the translated text can be thought of as the starting stage of the actual process (Martinez, 2004, p. 3). A long and complicated process ensues after the text translation, where the translator might also need to synchronize the text according to lip movement and proofread the text as well.

Ensuring the quality of dubbing doesn't include the spatial constraints of subtitling, but it does introduce a different layer of complexity which is synchrony (lip synchrony, kinetic synchrony, isochrony). Synchronizing is a process of coordinating the target language translation with the body language and lip movements of the characters appearing on screen. Since this aspect of dubbing is quite likely the most important, it is unsurprising that synchrony is the main focus of research in dubbing. Just as the quality of subtitles depends on the quality of the translated text; the quality of dubbing depends on the quality of synchrony of the audiovisual product (Conde Ruano, 2019, p. 134-35). A lot of professionals, technical equipment (both hardware and software) and a recording studio are needed to complete the process of dubbing, which means it takes a lot more money and time compared to subtitling. The following paragraphs will elaborate this lengthy process and the concept of quality control involved in it.

4.1. The Process of Dubbing

The process of dubbing is made up of four closely connected stages which are done in an already set order. If one of the stages is faced with an obstacle or a delay, it is quite likely that other phases will be affected as well, and since a large number of different individuals are working on the project problems are quite common. The major aspects of dubbing are very similar everywhere in the world, but some small changes may be implemented in different countries and/or studios.

Prior to production, a client (television station, broadcaster, streaming service) will make a request to a dubbing studio by sending a copy of the audiovisual product to them. Usually, both video and the script are sent along with a detailed list of instructions (e.g., voice actor suggestions, should songs be dubbed, etc.) to make the translation process easier. Those materials are then given to a translator, in most cases a freelancer, who uses them to produce a translated text. Sometimes the sent script isn't completely accurate or isn't sent at all in which the translator will use only the film to provide the dubbing studio with a translation. When the translated text is completed, it may sometimes be proofread by an employee of the client. Some broadcasters find this necessary, while others might skip proofreading entirely (Martinez, 2004, p. 3-4).

The following stage involves synchronizing the translated text with lip movements and body language of the original video. The synchronies taken into consideration are usually the following three types: lip synchrony (matching the lip movement and words which are uttered), kinetic synchrony (matching body movements with the audio track) and isochrony (establishing the duration of pauses and utterances) (Conde Ruano, 2019, 134). In certain cases, synchronization is done by the translator or the proofreader, although it is more common for an actor or the dubbing director to complete such a task. Quality synchronizations do not drift far in terms of meaning from the source text.

The production team receives the text after its synchronization and edits it so it is ready for the final stage of dubbing. Editing means cutting the synchronized text into takes (job usually done by a production assistant), which will be acted out during dubbing. The production team also organizes a schedule for the dubbing sessions, when each scene will be acted out and which actors need to be present for it. This is generally done in such a way that will take the least amount of time, and as a result reduce the cost of production.

After all of the mentioned phases have been completed, the dubbing begins. Voice actors perform the prepared material divided into takes while the dubbing director is present and guides them through the process, giving detailed instructions according to the client's demands. Of course, during voice recording all mistakes must be avoided, especially mispronunciation and content mistakes. Final editing carefully coordinates the recorded audio track and the moving image, which marks the end of the dubbing process. The audiovisual content is then ready for the audience to watch it.

4.2. Audience Reception and Quality Control in Dubbing

Before delving into the concept of quality control carried out by dubbing professionals, it is important to acknowledge the importance of audience reception, which spans through all modes of translation, not only dubbing. However, dubbed audiovisual products may involve certain errors in synchrony, and such mistakes can have a much bigger impact on the viewing experience of the audience. Even though quality control of professionals is an important aspect in translation of audiovisual products, the sole reason it is carried out is to ensure the translated product is received positively by its viewers.

Through observation it can be concluded that viewers focus mostly on voice acting and of course the visual synchrony involved in dubbing. To receive further and more specific insights into audience reception, additional research should be conducted on multiple levels. Firstly, the audience's preferences and needs should be explored in order to develop appropriate translation solutions. Secondly, information should be gathered through research of how viewers perceive different semiotic layers of audiovisual products. Very little progress has been made in terms of research in the field of audience reception in dubbing, so before drawing any final conclusions, additional surveys need to be carried out.

When it comes to quality control in dubbing, research is also quite sparse although some parameters have been proposed for measuring quality. The aspects of quality that should be checked include the consideration of body and lip movements, coherence between image and text, voice acting, technical standards, following the original script, etc. In 2019 a paper was published by Conde Ruano containing results of an experiment conducted on the topic of quality control in dubbing. The goal of the research was to observe which aspects are perceived as vital in determining the quality of a dubbed product by providing excerpts (some with video, some

with text-only) for subjects (second- and fourth-year university students in translation studies) to carry out quality control on. The excerpts contained various examples of errors in synchronization with a control excerpt which contained little to no errors (Conde Ruano, 2019, p. 135-37). Some students were given scenes including video for evaluation while the others only evaluated the written script. This was done in order to determine whether absence of video material will significantly influence quality ratings, since synchrony accuracy is easily observable while watching the product instead of only reading the script.

First hypothesis of the project confirmed that those excerpts that contained observable errors of synchrony are perceived to be of a lower quality than those not containing such errors. Although synchrony errors did affect the rating, the difference was not huge. Furthermore, it was first assumed that isochrony errors (mistakes involving duration between pauses and utterances) would influence quality rating the most, but in reality the scenes which lacked lip synchrony seem to be the most problematic.

Research also concluded that the presence of video during evaluation did not greatly influence the level of demand in terms of synchrony. Defying the expectations, the dubbed versions containing slight synchrony errors seem to be rated slightly better than text-only versions, possibly because of the more elements which are involved in it, and because of the satisfactory performances put in by the voice actors (Conde Ruano, 2019, p. 146-47).

In conclusion, it seems that dubbing and its quality control reach far beyond just the linguistic level. Aspects such as voice acting, tone and the entire soundtrack of the product appear to conceal the errors made during translation and adaptation. With progress in research, strides are being made in uncovering the mystery elements of dubbing.

5. Conclusion

Because of the ever-increasing volume of production of AV products, there is an increase in the need for AVT as well. Many modes of AVT were designed in order to make various content accessible to as many 'non-source' language speakers as possible, but also to those who are deaf or hard of hearing. Alongside two of the most popular modes of AVT, which are subbing and dubbing, several of less popular and less researched modes have been explored in this paper (namely video game localization, surtitling, voiceover and IRSP). Although all of these modes share certain aspects (text translation, spatial and/or temporal constraints), each of them has a unique process which requires more than just translation skills and it often takes several professionals to complete all parts of the process. Depending on the AVT mode a whole team can be involved in the process (when it comes to dubbing, video game localization and voiceover) or only one or two people can be involved (subbing, surtitling, IRSP).

However, all of the processes require a translator who usually has a much bigger task than just translating a text. During AVT, a translator might also need to synchronize the translated text with the image of the AV product (subbing, localization), cue the captions during a live performance or speech (surtitling, IRSP) and often proofread and/or test the final product for the purposes of quality control. Because all of these AVT processes are heavily reliant on technology, translators must be familiar and comfortable with the appropriate hardware and software required for process completion. For example, when producing subtitles, translators have to be familiar with software which allows them to timecode subtitles (set the exact time of their appearance and disappearance) and adjust their visual appearance according to quality guidelines. Technological advancements have made such tasks much easier and will continue to do so with constant improvements to both hardware and software.

Another interesting aspect of AVT often brought up is its polysemiotic nature. Unlike only written texts, AV products incorporate multiple semiotic layers into a unified whole. Non-interactive products (films, TV shows, operas, etc.) include five different semiotic layers, while interactive media (such as video games) have as many as eight semiotic layers. This is what makes AVT so unique and difficult. Such creative forms of media also demand creative modes of translation.

Finally, one of the central topics in AVT and translation studies in general is quality control. Quality control is an important part of AVT and is meant to be conducted thoroughly

after every translation according to professional quality standards. Many countries offer guidelines for quality control in translation to make the job simpler for translators. These guidelines usually consist of linguistic and technical aspects which should be fulfilled. Although translators do follow these guidelines, their work is supposed to undergo quality control performed by another translator (who is preferably very experienced). However, this part of the process might sometimes be partially or entirely skipped in order to reduce cost or save time, depending on the client's instructions. Such a decision may result in overlooked errors, hence a translation of lesser quality. Because of this, it is always preferable to employ quality controllers.

In conclusion, since a large part of academic works published on the topic of AVT focuses on dubbing and subtitling, they are used as a framework for furthering the research of less popular modes of AVT. This includes establishing quality guidelines and implementing improved systems for measuring the accuracy of translated products. One of the points for further research is conducting surveys on audience members to firmly establish the viewers' preferences, since audience reception is one of the key aspects of every product.

6. Works Cited

Bernal-Merino, Miguel. “Key Concepts in Game Localisation Quality.” Bogucki and Deckert, pp. 297-314.

Bogucki, Łukasz and Mikołaj Deckert, editors. The Palgrave Handbook of Audiovisual Translation and Media Accessibility. Palgrave, 2020.

Carrillo Darancet, Juan M. “The Drama of Surtitling: Ever-Changing Translation on Stage.” Bogucki and Deckert, pp. 173-88.

Chiaro, Delia. “Issues in Audiovisual Translation.” The Routledge Companion to Translation Studies, edited by Jeremy Munday, Routledge, 2009, pp. 141-65.

Conde Ruano, Jose T. “Dischronies and the Absence of Image in the Evaluation of Dubbing.” Quality Assurance and Assessment Practices in Translation and Interpreting, edited by Elsa Huertas-Barros, Sonia Vandepitte and Emilia Iglesias-Fernandez, IGI Global, 2019, pp. 133-54.

Davitti, E. & Sandrelli, A. (2020). Embracing the Complexity: A Pilot Study on Interlingual Respeaking. Journal of Audiovisual Translation, 3(2),103–139.

Dawson, H. (2019). Feasibility, quality and assessment of interlingual live subtitling: A pilot study. Journal of Audiovisual Translation, 2(2),36–56.

Dawson, H. (2020). A research-informed training course for interlingual respeaking. Journal of Audiovisual Translation, 3(2), 204–225.

de Linde, Zoe (1995). ‘Read my lips’ subtitling principles, practices, and problems, Perspectives: Studies in Translatology, 3:1, 9-20

Diaz-Cintas, Jorge. “The Name and Nature of Subtitling.” Bogucki and Deckert, pp. 149-72.

“Guidelines for Quality Subtitling in Croatia.” *Audiovisual Translators Europe*, 2022. Standards / Guidelines – AUDIOVISUAL TRANSLATORS EUROPE (avteurope.eu)

"Isochrony." *Definitions.net.* STANDS4 LLC, 2022. <https://www.definitions.net/definition/ISOCHRONY>.

Martinez, Xenia. “Film Dubbing, its Process and Translation.” Topics in Audiovisual Translation, edited by Pilar Orero, John Benjamins Publishing Company, 2004, pp. 3-8.

Matamala, Anna. “Voice-over: Practice, Research and Future Prospects.” Perez-Gonzalez, pp. 64-81.

Nikolić, K. (2021). Quality Control of Subtitles: A Study of Subtitlers, Proofreaders, Quality Controllers, LSPs and Broadcasters/Streaming Services. Journal of Audiovisual Translation, 4(3),66–88.

Perez-Gonzalez, Luis, editor. The Routledge Handbook of Audiovisual Translation. Routledge, 2019.

Remael, Aline and Nina Reviers. “Multimodality and Audiovisual Translation: Cohesion in Accessible Films.” Perez-Gonzalez, pp. 260-80.

Robert, Isabelle S. and Aline Remael.. "Quality Control in the Subtitling Industry: An Exploratory Survey Study." Meta 613 (2016): 578–605.

Romero-Fresco, P., & Martínez, J. (2015). Accuracy rate in live subtitling–the NER Model. In J. Díaz-Cintas, & R. Baños Piñero (Eds.), Audiovisual translation in a global context. Mapping an ever-changing landscape (pp. 28–50). London: Palgrave.

Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR Model. Linguistica Antverpiensia,16,149–167.

Secară, Alina. “Surtitling and Captioning for Theatre and Opera.” Perez-Gonzalez, pp. 130-44.

"Semiotic." *Definitions.net.* STANDS4 LLC, 2022. <https://www.definitions.net/definition/semiotic>.

"Surtitles." *Definitions.net.* STANDS4 LLC, 2022. <https://www.definitions.net/definition/surtitles>.

Taylor, Christopher. “Multimodality and Audiovisual Translation.” Handbook of Translation Studies Volume 4, edited by Yves Gambier and Luc van Doorslaer, John Benjamins Publishing Company, 2013, pp. 98-104.